

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-099281  
(43)Date of publication of application : 07.04.2000

(51)Int.Cl. G06F 3/06

(21)Application number : 10-264286  
(22)Date of filing : 18.09.1998

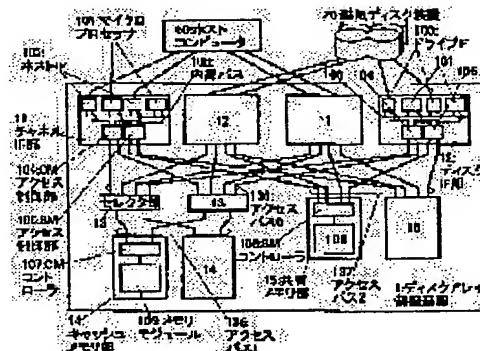
(71)Applicant : HITACHI LTD  
(72)Inventor : FUJIMOTO KAZUHISA  
TANAKA ATSUSHI  
FUJIBAYASHI AKIRA  
KANAI HIROKI  
MINOWA NOBUYUKI

## (54) DISK ARRAY CONTROLLER

(57) Abstract:

**PROBLEM TO BE SOLVED:** To improve throughput by making the style of a connection between plural channel interface and disk interface parts and a cache memory part different from the style of a connection with a shared memory part.

**SOLUTION:** A channel interface part 11, a disk interface part 12 and a cache memory part 14 are connected through a selector part 13. Besides, the channel interface part 11, disk interface part 12 and shared memory part 15 are directly connected without interposing the selector part 13. Thus the number of access paths to be directly connected to the cache memory part is reduced, throughput is improved, processing overhead at the selector part is eliminated and time for access to the shared memory part can be shortened.



## LEGAL STATUS

[Date of request for examination] 25.03.2003

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's

decision of rejection]

[Date of extinction of right]

Copyright (C): 1998,2003 Japan Patent Office

## \* NOTICES \*

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

## CLAIMS

## [Claim(s)]

[Claim 1] Two or more channel-interface sections which have an interface with a host computer. An interface with a magnetic disk unit. It is the disk array control unit equipped with the above, and is characterized by the connection form between two or more aforementioned channel-interface sections and two or more aforementioned disk interface sections, and the aforementioned cache memory section differing from the connection form between two or more aforementioned channel-interface sections and two or more aforementioned disk interface sections, and the aforementioned shared memory section.

[Claim 2] It is the disk array control unit according to claim 1 which it connects through the selector section between two or more aforementioned channel-interface sections and two or more aforementioned disk interface sections, and the aforementioned cache memory section, and is characterized by carrying out the direct file of between two or more aforementioned channel-interface sections and two or more aforementioned disk interface sections, and the aforementioned shared memory section without the selector section, respectively.

[Claim 3] Each is a disk array control unit according to claim 1 which it connects through the selector section between two or more aforementioned channel-interface sections and two or more aforementioned disk interface sections, and the aforementioned cache memory section, and is characterized by taking 1 to 1 \*\*\*\* between two or more aforementioned channel-interface sections and two or more aforementioned disk interface sections, and the aforementioned shared memory section.

[Claim 4] It is the disk array control unit according to claim 1 characterized by connecting through the selector section between two or more aforementioned channel-interface sections and two or more aforementioned disk interface sections, and the aforementioned cache memory section, and connecting two or more aforementioned channel-interface sections, two or more aforementioned disk interface sections, and the aforementioned shared memory section to a share bus.

[Claim 5] It is the disk array control unit according to claim 1 which the cross coupling network which used the switch connects between two or more aforementioned channel-interface sections and two or more aforementioned disk interface sections, and the aforementioned cache memory section, and is characterized by carrying out the direct file of between two or more aforementioned channel-interface sections and two or more aforementioned disk interface sections, and the aforementioned shared memory section, respectively.

[Claim 6] It is the disk array control unit according to claim 1 which the cross coupling network which used the switch connects between two or more aforementioned channel-interface sections and two or more aforementioned disk interface sections, and the aforementioned cache memory section, and is characterized by taking 1 to 1 \*\*\*\*, respectively between two or more aforementioned channel-interface sections and two or more aforementioned disk interface sections, and the aforementioned shared memory section.

[Claim 7] Two or more channel-interface sections which have an interface with a host computer. Two or more disk interface sections which have an interface with a magnetic disk unit. The cache memory section which stores temporarily the data by which are connected to two or more

aforementioned channel-interface sections and two or more aforementioned disk interface sections, and read/write is carried out to the aforementioned magnetic disk unit. The shared memory section which is connected to two or more aforementioned channel-interface sections and two or more aforementioned disk interface sections, and stores the control information about the data transfer between the channel-interface section and the disk interface section, and the aforementioned cache memory section. It is the disk array control unit equipped with the above, and the number of the access path connected to the aforementioned cache memory section is characterized by being fewer than the number of the access path connected to the aforementioned shared memory section.

[Claim 8] It has the selector section further. each channel-interface section and each disk interface section, and the aforementioned selector section An access path connects with 1 to 1, respectively. the aforementioned selector section and the aforementioned cache memory section The total of the access path which is connected by the access path and connects each channel-interface section and each disk interface section, and the aforementioned selector section More [ total / of the access path which connects the aforementioned selector section and the aforementioned cache memory section ] It is the disk array control unit according to claim 7 characterized by the access path connecting with 1 to 1, respectively between each channel-interface section and each disk interface section, and the aforementioned shared memory section.

[Claim 9] Two or more interface sections with a host computer. Two or more interface sections with a magnetic disk unit. The cache memory section which stores the data of a magnetic disk unit temporarily and which became independent physically. Having the shared memory section which stores control information and which became independent physically, each interface section with the aforementioned host computer and each interface section with the aforementioned magnetic disk unit are a microprocessor, respectively. The cache memory control section which controls access to the aforementioned cache memory section, and the share memory access control section which controls access to the aforementioned shared memory section. It is the disk array control unit equipped with the above, and the access path connects through the selector section between each cache memory access control section and the aforementioned cache memory section, and it is characterized by the direct file being carried out by the access path, respectively between each share memory access control section and the aforementioned shared memory section.

[Claim 10] Two or more interface sections with a host computer. Two or more interface sections with a magnetic disk unit. The cache memory section which stores the data of a magnetic disk unit temporarily and which became independent physically. Having the shared memory section which stores control information and which became independent physically, each interface section with the aforementioned host computer and each interface section with the aforementioned magnetic disk unit are a microprocessor, respectively. The cache memory control section which controls access to the aforementioned cache memory section, and the share memory access control section which controls access to the aforementioned shared memory section. It is the disk array control unit equipped with the above, and the cross coupling network which used the switch connects between each cache memory access control section and the aforementioned cache memory section, and it is characterized by the direct file being carried out by the access path, respectively between each share memory access control section and the aforementioned shared memory section.

[Claim 11] A disk array control unit given in any of the claim 9 characterized by carrying out bandwidth of the access path between each cache memory control section and the aforementioned cache memory section to more than the double precision of all the greatest bandwidth between the interface sections of the aforementioned host computer and the aforementioned host computer, or a claim 10 they are.

[Claim 12] A disk array control unit given in any of the claim 1 characterized by having had two or more aforementioned shared memory sections and aforementioned cache memory sections, and having doubled by two or more of these shared memory sections and two or more aforementioned cache memories, or a claim 11 they are.

---

[Translation done.]

## \* NOTICES \*

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.

2. \*\*\* shows the word which can not be translated.

3. In the drawings, any words are not translated.

## DETAILED DESCRIPTION

## [Detailed Description of the Invention]

[0001] [The technical field to which invention belongs] this invention relates to the control unit of the disk array equipment which stores data in two or more magnetic disk units.

[0002] [Description of the Prior Art] Compared with the I/O performance of the primary storage of the computer which uses a semiconductor memory as a storage, the disk-subsystem (henceforth "subsystem") I/O performance which uses a magnetic disk as a storage is small about 3-4 times, and efforts to raise contracting this difference from the former, i.e., the I/O performance of a subsystem, are made. As one method for raising the I/O performance of a subsystem which stores data in two or more magnetic disk units, and the system which stores

subsystem is constituted from two or more magnetic disk array which is called so-called disk array. Two or more channel IF sections 11 which perform data transfer between a host computer 50 and the disk array control unit 2, Two or more disk IF sections 12 which perform data transfer between a magnetic disk unit 20 and the disk array control unit 2. The cache memory section 15 which stores the data of a magnetic disk unit 20 temporarily, it has the shared memory section 14 about the data transfer control between the channel IF section and the disk IF section, and the cache memory section 15 have array control unit 2. The cache memory section 14 and the shared memory section 12. In this disk accessible composition from all the channel IF sections 11 and the disk IF section 12, and array, it connects with 1 to 1 between the channel IF section 11 and the disk IF section 12, the shared memory section 15 and between the channel IF section 14 and the disk IF section 12, and the cache memory section 14. Hereafter, such a topology is called star connection.

[0004] The channel IF section 11 has the microprocessor (not shown) which controls the interface for connecting with a host computer 50, and the I/O over a magnetic disk unit 20. Moreover, the disk IF section 12 which perform data transfer between a magnetic

control units 3. Two or more disk IF sections 14 which perform data transfer between the channel IF section 11 and the disk array control unit 2. The cache memory section 15 which stores the data of a magnetic disk unit 20 temporarily, it has the shared memory section 14 about the data transfer control between the disk array control information (for example, information about the data transfer between the channel IF section 11 and the disk IF section 12, and the cache memory section 14) about the data transfer control between the channel IF section 11 and the disk IF section 12, and the cache memory section 14. Hereafter, such a topology is called star connection.

[0005] Drawing 3 shows the composition of other conventional disk arrays. A host computer 50 and two or more channel IF sections 11 which perform data transfer between a magnetic control units 3. Two or more disk IF sections 14 which perform data transfer between the channel IF section 11 and the disk array control unit 2. The cache memory section 15 which stores the data of a magnetic disk unit 20 temporarily, it has the shared memory section 14 about the data transfer control between the disk array control information (for example, information about the data transfer between the channel IF section 11 and the disk IF section 12, and the cache memory section 14) about the data transfer control between the channel IF section 11 and the disk IF section 12, and the cache memory section 14. Hereafter, such a topology is called star connection.

[Problem(s) to be Solved by the Invention] In order to make architecture of a disk array scalable, according to the number of channels with a required host computer, it is necessary to extend the disk IF section and to extend the channel IF section in a disk array control unit according to the disk capacity (the number of logical volumes) linked to a disk controller. However, since the transfer capability of the once mounted share bus cannot be changed according to extension of the channel IF section and the disk IF section in the disk array control unit of the share bus the channel IF section and the disk IF section in the view 3, it is difficult to deal with extension of the channel IF section and the disk IF section flexibly.

[0007] moreover, in the disk array control unit of the share bus connection form shown in the view 3 The microprocessor which controls the I/O over the host computer prepared in the channel IF section which performs data transfer between a host computer and a disk array control unit. And when a highly efficient processor is used for the microprocessor which performs data transfer the I/O over the magnetic disk unit prepared in the disk IF section which performs data transfer between a magnetic disk unit and a disk array control unit Compared with the performance of these processors, the transfer capability of a share bus becomes a bottleneck and it becomes difficult to follow improvement in the speed of a processor.

[0008] Furthermore, when an obstacle occurs in the share bus connection form shown in the view 3 in which channel IF section (or the disk IF section) of two or more channel IF sections (or two or more disk IF sections) connected to the share bus, it is difficult to specify the channel IF section (or the disk IF section) which the obstacle generated.

[0009] It is possible to make an internal path performance increase according to the performance of the processor which the channel IF section and the disk IF section extend or use on the other hand, since an internal path performance increase in proportion to the number of access paths connected to the shared memory section or the cache memory section in the disk array control unit of the star connection form shown in the view 2. Moreover, in the disk array control unit of the star connection form and the disk IF section, and the shared between the channel IF section and the disk IF section, and the shared between the channel IF section and the disk IF star connection of between the channel IF section (or the disk IF memory sections is carried out, it is also easy to specify the channel IF section (or the disk IF section) which the obstacle generated.

[0010] In the disk array control unit of star connection form, when the number of the channel IF section carried or the disk IF sections is increased, the number of access paths between the channel IF section and the disk IF section, and the cache memory section and between the channel IF section and the disk IF section, and a shared memory will also increase. To

[0011] However, the amount of one control information stored in a shared memory in the disk controller connected to a

[0012] the amount of data of one shared memory in the disk controller connected to a

[0013] one control information stored in a shared memory being about (for example, 2 K bytes) several K

[0014] bytes is about (for example, 4 bytes) several bytes. Moreover, one control information

[0015] amount of data transmitted between the channel IF section and the disk IF section, and the cache memory section is increased, and it is necessary to raise an internal path performance.

[0016] Moreover, the amount of data stored in a cache memory is quite larger than a shared memory is increased, and it is necessary to raise an internal path performance.

[0017] If an example is given, the amount of data of one data stored in a shared memory [0018] However, the amount of data of one control information stored in a shared memory being about (for example, 2 K bytes) several K bytes is about (for example, 4 bytes) several bytes. Therefore, since many [ compared with the system to one data stored in a cache memory being about (for example, 2 K bytes) several K bytes is about (for example, 4 bytes) several bytes. Moreover, one control information

[0019] amount of data transmitted between the channel IF section and the disk IF section, and the cache memory section / quite ]. the amount of data width of face of

[0020] shared memory section and the cache memory section needs to take the data width of face of

[0021] the cache memory section, and the cache memory section and the cache memory

[0022] and the disk IF section, and the access path between the channel IF section and the disk IF section

[0023] the access path between the channel IF section and the disk IF section, and the former access path

[0024] section larger than the data width of face of the access path. For example, the former access path

[0025] and the disk IF section, and the shared memory section. For example, the former access path

[0026] consists of buses of 16-bit width of face, and the latter consists of buses of 4-bit width of face.

[0027] Hereafter, such a topology is called share bus connection form.

2003/06/10

2003/06/10

Therefore, if the number of the access path between the channel IF section and the disk IF section, and the cache memory section is increased, the problem that the numbers of pins of LSJ of the cache memory section which connects those access paths run short will arise. Moreover, in order to shorten the response time to the host computer of a disk array control unit, it is also required to shorten the access time to the control information stored in the shared memory section as much as possible.

[0012] Then, the purpose of this invention is to offer the subsystem in consideration of the property of the data stored in a cache memory and a shared memory, and the access property to such memory for a throughput to be high and using a disk array control unit and its section [0013] More specifically, the access path between the channel IF section and the disk IF section and a shared memory has the access time in offering a short disk array control unit and the subsystem using it highly [the access path between the channel IF section and the disk IF section, and the cache memory section has / the purpose of this invention / a high throughput and / a throughput ].

[0014] Means for Solving the Problem] Two or more channel-interface sections in which the above-mentioned purpose has an interface with a host computer. Two or more disk interface sections which have an interface with a magnetic disk unit. The cache memory section which stores temporally the data by which read/write is carried out to the aforementioned magnetic disk unit, it has the shared memory section which stores the control information about the data transfer between the channel-interface section and the disk interface section, and the aforementioned cache memory section. Each channel-interface section performs data transfer between an interface with the aforementioned host computer, and the aforementioned cache memory section. In the disk array control unit which performs data transfer section, each disk interface section and the aforementioned cache memory section, each channel-interface section and the aforementioned channel-interface section, the connection form between two or more aforementioned channel-interface sections and two or more aforementioned disk interface sections, and the aforementioned memory section. The connection form between two or more aforementioned cache memory sections and two or more aforementioned disk interface sections, and the aforementioned memory section. It is attained by the disk array control unit characterized by differing from the connection form between two or more aforementioned channel-interface sections and two or more aforementioned disk interface sections, and the aforementioned shared memory section. [0015] It connects through the selector section between two or more aforementioned channel-interface sections and two or more aforementioned disk interface sections, and the aforementioned cache memory section, and the direct file of between two or more aforementioned channel-interface sections and two or more aforementioned disk interface sections and two or more aforementioned shared memory section is carried out without the selector

are clarified with the column and drawing of an operator. e. [0019] [Embodiments of the Invention] Hereafter, the example of this invention is explained using a drawing.

[0020] One example of this invention is shown in [example 1] drawing 1.

[0021] The disk array control unit 1 has the two interface sections (channel IF section) 11 with a

[0027] Two access paths 2:137 are connected to SM access-control section 105, and the access path 2:137 is connected to two different SM controllers, respectively. Therefore, every channel IF sections 11 and two disk IF sections 12. In this example, although one access path 2:137 was connected between one SM access-control section 105 to one SM controller 108, since the access root from one SM access-control section 105 to two obstacle-proof nature can be raised.

http://192.168.1.100/cgi-bin/tran\_web.cgi\_ejje

2000/06/10

[0030] The big feature of this example is in the point that the connection form between the channel IF section 11 and the disk IF section 12, and the cache memory section 14 differs from the connection form between the channel IF section 11 and the disk IF section 12, and the shared memory section 15. The reason made such composition is explained using drawing 1 and drawing 2. In the disk array control unit 2 of the star connection form shown in drawing 2, when the number of the channel IF section 11 carried in the disk array control unit 2 or the disk IF sections 12 is increased, the number of access paths between the channel IF section 11 and the disk IF section 12, and the shared memory section 15 will also increase. Moreover, in the disk IF section 12, and the cache memory section 14 and between the channel IF section 11 and the disk IF section 12, and the shared memory section 15, and to raise an internal path performance.

[0031] However, the amount of one data stored in the cache memory section 14 is quite larger than the amount of data of one control information stored in the shared memory section 15. If an example is given, one control information stored in the shared memory section 15 in the disk controller connected to a mainframe to one data stored in the cache memory section 14 being about (for example, 2 Kbytes) several Kbytes will be about (for example, 4 bytes) several bytes. Moreover, one control information stored in the shared memory section 15 in the disk controller connected to the host computer of an open system to one data stored in the cache memory section 14 being about (for example, 64 bytes) dozens of bytes is about (for example, 4 bytes) several bytes. Therefore, the amount of data transmitted between the channel IF section 11 and the disk IF section 12, and the cache memory section 14 since many [compared with the former access path between the channel IF section 11 and the disk IF section 12, the amount of data transmitted between the channel IF section 11 and the disk IF section 12, and the shared memory section 15 / quite ] it is necessary to take the data width of face of the access path between the channel IF section 11 and the disk IF section 12, and the cache memory section 14 larger than the data width of face of the access path between the channel IF section 11 and the disk IF section 12, and the shared memory section 15. For example, the number of pins of LSI of the cache memory controller in the cache memory section 14, (not shown [the cache memory controller] in drawing 2), or the package which mounts the cache memory section 14, the problem that the number of pins of a connector is insufficient will arise. Then, in this example, as shown in drawing 1, the number of access paths by which a direct file is carried out to the cache memory section 14 is cut down by connecting between the selector IF section 11 and the disk IF section 12, and the cache memory sections 14 through the selector section 13.

[0032] Since the data length of one control information stored in the shared memory section 13 is quite small compared with the data length of one data stored in the cache memory section 14 as mentioned above on the other hand, data width of face of the access path between the channel IF section 11 and the disk IF section 12, and the shared memory section 15 can be carried out to below the half of the data width of face of the access path between the channel IF section 11 and the disk IF section 12, and the cache memory section 14. Therefore, even if it increases the number of access paths to the shared memory section 15, it is rare for problems, such as a shortage of the number of pins of LSI of the shared memory memory controller (not shown [the share memory controller] in drawing 2) of shared memory circles, to arise.

[0033] Moreover, in order to shorten the response time to the host computer 50 of the disk array control unit 1, it is necessary to shorten the access time to the control information stored in the shared memory section 15 as much as possible. However, like [between CM access-control section 104 shown in drawing 1, and the CM controller 107], if between SM access-control section 105 and the SM controllers 108 is connected through the selector section, the access time to the control information stored in the shared memory section 15 cannot be shortened by the overhead of processing in the selector section.

[0034] Then, in this example, while preparing two or more access paths between the channel IF section 11 and the disk IF section 12, and the shared memory section 15 and raising a throughput by carrying out the direct file of between the channel IF section 11 and the disk IF section 12, and the shared memory sections 15 without the selector section, the processing overhead in the selector section is abolished and the access time from the disk, IF section 12 to the channel IF section 11 and the shared memory section 15 is made shortening.

[0035] In addition, at this example, although the selector section 13, the cache memory section 14, and the shared memory section 15 are doubled from a viewpoint of improvement in obstacle-proof nature, respectively, even if it does not double these, it cannot be overemphasized that an above-mentioned effect can be acquired.

[0036] Drawing 4 shows the composition in CM access-control section 104, CM access-control section 104 has the path IF 301 of a selector 302, the packet buffer 303 which stores the addresses, a command, and data temporarily, and the access path 0:135 connected with the selector section 13, the error-checking section 300 of data, and the data transfer control section 310. Two ports of a selector 302 are connected to a host IF 102 or drive IF 103 by the data line 210. Moreover, other two ports of a selector 302 are connected to path IF 301. A path IF 301 is connected to the selector section 13 by the access path 0:135. It connects with a host IF 102 or drive IF 103 by the control line 1:211, and the data transfer control section 310 is connected to the data transfer control section 315 in the selector section 13 by the control line 2:212. Moreover, the data transfer control section 310 performs the Arbitration of a host IF 102 or the access demand from drive IF 103 by the arbiter 308, and changes a selector 302. Drawing 5 shows the composition in the selector section 13. Four paths IF 301 with the access path 0:135 to which the selector section 13 is connected with the channel IF section 11 and the disk IF section 12. Two paths IF 301 with the access path 1:136 connected with the CM controller 107. The selector 306 which connects between both mutually, and a packet buffer 303. It has the error-checking section 300 of data, the address command (adr, cmd) analysis section 305 which analyzes the address and the command which were sent out from CM access-control section 104, and the data transfer control section 315. It connects with the data transfer control section 310 in CM access-control section 104 by the control line 2:212, and the data transfer control section 315 is connected to the data transfer control section 315 in the CM controller 107 by the control line 3:213. Moreover, by the arbiter 308, the data transfer control section 315 performs the Arbitration of the access demand from four access paths 0:135 analyzed in adr and the cmd analysis section 305, and changes a selector 306. A packet buffer 303 buffers some or all of data that is transmitted, in order to absorb the speed difference, when a difference is in a data transfer rate with the path by the side of an access path 0:135, and the path by the side of an access path 1:136.

[0038] adr and the cmd analysis section 305 have the buffer which stores the address and a command, the adr extraction section, and the cmd extraction section (not shown), four access paths 0:135 connected to CM access-control section 104, in adr and the cmd analysis section 305 — the address and a command are stored in the buffer which was alike, respectively and was assigned one [at a time]. In the adr extraction section and the cmd extraction section, the kind of the CM controller 107 to access and access is deduced, and it sends out to the arbiter 308 in the data transfer control section 315.

[0039] Drawing 7 shows the composition in the cache memory section 14. The cache memory section 14 has the CM controller 107 and a memory module 109. The CM controller 107 which are connected with the selector paths IF 301 with an access path 1:136 and selectors 304 which are connected with the selector section 13, the packet buffer 303 which stores data temporarily, the error-checking section 300 of data, the memory control section 307 which controls access to a memory module 109, adr and the cmd analysis section 305 which analyze the address and the command which were sent out from CM access-control section 104, and the data transfer control section 315. The data transfer control section 315 is connected to the data transfer control section 315 in the selector section 13 by the control line 3:213. Moreover, by the arbiter 308, the data transfer control section 315 performs the Arbitration of the access demand from two access paths 1:136 section 315 and changes a selector 304.

[0040] adr and the cmd analysis section 305 have a buffer, the adr extraction section, and the cmd extraction section (not shown), two access paths 1:136 connected to the CM controller 107 in adr and the cmd analysis section 305 — the address and a command are stored in the buffer which was alike, respectively and was assigned one [at a time]. In the adr extraction section and the cmd extraction section, the address of the memory to access and the kind of access are deduced and it sends out to the memory control section 307. Moreover, the access demand from two access paths 1:136 is sent out to the arbiter 308 in the data transfer control section 315. [0041] Next, the procedure at the time of access to the cache memory section 14 is described. When accessing to the cache memory section 14, a microprocessor 101 directs the access start to the cache memory section 14 to a host IF 102 or drive IF 103. [0042] The host IF 102 who received directions of an access start, or drive IF 103 sends out the signal which shows an access start by the control line 1:21 to the data transfer control section [0043] CM access-control section 104 stores in a packet buffer 303 the address, a command, and data (at the time of the writing of data). The data transfer control section 104 sends out through the data line 210, a command, and data (at the time of the writing of data) are sent out through the data line 210. [0044] Drawing 9 shows the flow of access to the CM controller 107 in the case of writing data in the cache memory section 14 from CM access-control section 104. The data transfer control section 310 in CM access-control section 104 will take out the signal (REQ) which shows an access start to the data transfer control section 315 in the selector section 13 by the control section 310 performs an Arbitration (Step 501), changes a selector 302. [0045] If a REQ signal is received from CM access-control section 104, the data transfer control section 13 will receive the address and the command which are sent through an access path 0:135 next, and will perform an Arbitration based on the access demand section 315 in the selector section 13 (Step 503). If the right of connection to an analyzed in adr and the cmd analysis section 305 (Step 503), the data transfer control section 315 — a selector 306 — changing (Step 504) — the signal (ACK) which shows that the right of access path 1:136 is acquired as a result of an Arbitration — the data transfer control section 315 — a selector 306 — changing (Step 504) — the signal (ACK) which shows that the right of access path 1:136 was acquired by the control line 2:2/2 to the data transfer control section 315. Next, the data connection to an access path 1:136 is returned (Step 505) Next, the data control section 310 in CM access-control section 104 is returned (Step 505) Next, the data transfer control section 315 takes out the signal (REQ) which shows an access start to the data transfer control section 315 in the CM controller 107 by the control line 3:2/3 (Step 506). The address and a command are sent out (Step 507). [0046] If an ACK signal is received, CM access-control section 104 reads data from a packet buffer 303, and sends them out to an access path 0:135 to an access path 0:135 to an access path 1:136 through a selector 301 and a selector 306 (Step 509). [0047] If a REQ signal is received by the control line 3:2/3, the data transfer control section 315 — a selector 306 — changing (Step 504) — the signal (ACK) which shows that the right of access path 1:136 is acquired as a result of an Arbitration — the data transfer control section 315 — a selector 306 — changing (Step 504) — the signal (ACK) which shows that the right of access path 1:136 are stored in a packet buffer 303. If the access privilege to a memory module 109 is acquired as a result of an Arbitration, the control information of memory module 109 is read from a packet buffer 303, and pretreatment for memory access will be sent out to the memory control section 307, and pretreatment for memory access will be performed (Step 510). Next, data are read from a packet buffer 303 and it writes in a memory module 109 through a selector 304 (Step 511). [0048] After access to a memory module 109 is completed, after treatment of memory access and the status (STATUS) which shows an access situation in the data transfer control section 315 is generated (Step 512). Next, the status is sent out to CM access-control section 310 through the selector section 13 (Step 513). The data transfer control section 310 in CM access-control section 310 in CM access-control section 310 turns off the REQ signal to the CM controller 107, if the status is

tums off the REQ signal to the selector section 13, if the status is 1. The CM access-transfer control section 315 in the selector section 13 turns off the ACK signal to CM access-control section 104, if OFF of the REQ signal from CM access-control section 104 is checked (Step 516).

[0049] The data transfer control section 310 in CM access-control section 104 will report the end of access to the cache memory section 14 to a host, if 102 or drive, if 103 by the control line 1:211, if the status is received.

[0050] The flow of access to the CM controller 107 in the case of reading data from the cache memory section 14 from CM access-control section 104 is the same as the case of the writing of data even with Steps 501-506 after Step 512.

[0051] CM access-control section 104 will go into the receiving waiting state of data here, if an error is received at Step 505.

[http://www4.ipdljpo.go.jp/cgi-bin/tran\\_web.cgi\\_ejje](http://www4.ipdljpo.go.jp/cgi-bin/tran_web.cgi_ejje)

memory module 109, and sends out data to an access path 2:137. path IF 301.

[0066] SM access-control section 105 sends out data to a microprocessor 101 through a selector 302 and the data line 220, if data are received through an access path 2:137. [0067] It is possible to carry two or more channel IF sections 12 which have two channels with a host computer 50 in the disk array control unit 1 of this example, and to connect those channels to a host computer different, respectively. In such a case, it is necessary to process all demands from each channel linked to the host computer in parallel.

[0068] By the way, in the disk array control unit 1, when reading data to a host computer 50, the data stored in the magnetic disk unit 20 are written in the cache memory section 14 through the disk IF section 12, and the data is read from the cache memory section 14, and is sent to a host computer 50 through the channel IF section 11. Moreover, when writing data in the disk array control unit 1 from a host computer 50, the data sent to the channel IF section 12 from the host computer 50 are written in the cache memory section 14, and the data is read from the cache memory section 14, and it writes in a magnetic disk unit 20 through the disk IF section 12. In order to generate the parity of data furthermore and to write in a magnetic disk unit 20, further two - three accesses are performed between the disk IF section 12 and the cache memory section 14.

[0069] Therefore, in order to process all demands from each channel linked to the host computer in parallel, let the throughput between the channel IF section 11 and the disk IF section 12, and the cache memory section 14 be a host computer 50 more than the double precision of the greatest throughput between the channel IF sections 11.

[0070] In this example, the bandwidth of the access path between CM access-control section 105 and a CM controller 107 and a

[http://www4.ipdl.jp/cgi-bin/tran\\_web.cgi\\_EU](http://www4.ipdl.jp/cgi-bin/tran_web.cgi_EU)

[http://www4.ipdl.jp/o.go.jp/cgi-bin/tran\\_web\\_cgi.cgi](http://www4.ipdl.jp/o.go.jp/cgi-bin/tran_web_cgi.cgi)

[0083] After access to a memory module 105 is completed, the data transfer performed and the status (STATUS) which shows an access situation in the data transfer control section 315 is generated (Steps 712a and 712b). Next, the status is sent out to CM access-control section 104 through the selector section 13 (Steps 713a and 713b). The data transfer control section 315 in the selector section 13 turns off the CM controller a and the transfer control section 315 in the selector section 13 turns off the CM controller b. Moreover, if the status is received (Steps 714a and 714b), the selector section 13 REQ signal of b:107, respectively, if the status is received from both the CM controller a and b:107, the selector section 13 continues them and sends them out to CM access-control section 104 turns off the REQ signal to the selector control section 310 in CM access-control section 104 turns off the data transfer control section 315 if section 13, if the two statuses are received (Step 715). The data transfer control section 104, if OFF of the selector section 13 turns off the ACK signal to CM access-control section 104 is checked (Step 716).

[0084] The data transfer control section 14 to a host JF 102 or drive JF 103 by the control end of access to the cache memory section 14 is received.

line 1:211, if the status is received.

[0085] When writing data in the cache memory section 14 doubled as mentioned above doubly, while performing write-in processing of one data of the data written in doubly, another data need to prevent being rewritten by the write request from other access paths. In this example, the to prevent being rewritten by the write request from other access paths. In this example, the directory of the cache memory section 14 is stored in the shared memory section 15, and before accessing the cache memory section 14, the bit which shows under access to the directory stored in the shared memory section 15 is surely stood. Since only one access demand is simultaneously published by this by the same address in the cache memory section 14, while performing write-in processing of one data of the data written in doubly, another data can prevent being rewritten by the write request from other access paths.

[0086] The [example 3] disk array control unit 1 requires the function which copies data to another cache memory section 14 from a certain cache memory section 14, when two or more cache memory sections 14 are formed. This function is realizable with the procedure described below.

[0087] The host JF 102 who received directions of an access start, or drive JF 103 sends out signal which shows an access start by the control line 1:211 to the data transfer control section 14. The data transfer control section 104. With it, the two addresses and a command are sent or

through the data line 210. 1 of the two addresses and a command are the addresses and lead commands of a copied material, and another address and a command are the addresses and the light commands of a copy place. Here, the CM controller b is explained as a copy place CM controller a:107 copy-origin.

[0088] CM access-control section 104 stores in a packet buffer 303 the two addresses and the command which have been sent through the data line 210. The data transfer control section 310 performs an Arbitration, determines the royalty of a path If 301, and changes a selector 302. Drawing 12 shows the CM controller a from CM access-control section 104 in the case of copying data between the two cache memory sections, and the flow of access to b:107. The data transfer control section 310 in CM access-control section 104 will take out the signal (REQ) which shows an access right to the data transfer control section 315 in the selector section 13 by the control line 2:212, if the royalty of an access path 0:135 is determined by the arbitration (Step 11). Then, the address and two commands are sent out continuously (Step 12).

Arbitration (Step 303) . . . . .

802). [0050] If a REQ signal is received from CM access-control section 104, the data transfer control section 315 in the selector section 13 will receive the address and the command which are sent through an access path 0:135 next, and will perform an Arbitration based on the access demand analyzed in adr and the cmd analysis section 305 (Step 803). if both the CM controller a and the right of connection to two access paths 1:136 of b:107 are acquired as a result of an Arbitration — the signal — the data transfer control section 315 — a selector section 306 — changing (Step 804) — the signal (ACK) which shows that the right of connection to the CM controller a and b:107 both was acquired by the control line 2:12 to the data transfer control section 310 in CM access-control section 104 is returned (Step 805) Next, the data transfer control section 315 takes out the signal (REQ) which shows an access start by the control line 3:213 to the CM controller a and signal (REQ) which shows an access start by the control line 3:215 in b:107 (Steps 806a and 807a).

the data transfer control section 104, a controller a and b:107 (Steps 307a and 307b). Command are sent out to the CM controller a and b:107 (Steps 307a and 307b). [0091] CM access-control section 104 will go into the receiving waiting state of the status which tells the end of access, if an ACK signal is received.

[0092] If a REQ signal is received by the control line 3:213 the data transfer control section 315 are sent through an access path 1:136 next, will perform an Arbitration based on the access demand analyzed in adr and the cmd analysis section 305 (Step 808), and will change a selector 304. If the access privilege to a memory module 109 is acquired as a result of an Arbitration, the control information of memory will be sent out to the memory control section 307, and pretreatment for memory access will be performed (Step 809). Next, data are read from a memory module 109 (Step 810), and it sends out to an access path 1:136 through a selector 304.

"107 through an

2003/06/10

821). Moreover, if the status is received from both two CM controllers a and b:107, the selector section 13 continues them and sends them out to CM access-control section (Step 820). The data transfer control section 310 in CM access-control section 104 turns off the REQ signal to the selector section 13, if the two statuses are received (Step 822). The data transfer control section 315 in the selector section 13 turns off the ACK signal to CM access-control section 104, if OFF of the REQ signal from CM access-control section 104 is checked (Step 823). [0087] The data transfer control section 310 in CM access-control section 104 will report the end of access to the cache memory section 14 to a host IF 102 or drive IF 103 by the control line 1:211, if the status is received.

[0088] The disk array control unit 1 also requires the function which copies data to another address from a certain address in the one cache memory section 14 again. [0089] This function is realizable by being in the procedure at the time of the writing of the data shown by drawing 9 by performing processing to the memory module 109 of Step 511 in which store in the packet buffer 303 in the CM controller 107 by reading data from a memory module 109, instead of light access, and the data is continuously written in a memory module 109. [0100] In [Example 4] drawing 1, it becomes possible to raise the reliability of data by writing in the same data as two fields which doubled all or a part of memory storage, and were doubled between the two shared memory sections 15 at the time of the writing of the data to the shared memory section 15.

[0101] The procedure in the case of writing data in the two doubled shared memory sections 1 is as follows.

[0102] A microprocessor [101] sends out the signal which shows the SM access-control section 105. With it, the line 5/221 to the data transfer control section 310 in SM access-control section 105. When two addresses, a command, and one data are sent out through the data line 220, [0103] SM access-control section 105 stores in a packet buffer 303 the two addresses and the command which have been sent through the data line 220, and one data. The data transfer control section 310 performs an Arbitration, determines the royalty of a path [F:301, and changes a selector 302.

[0104] Drawing 13 shows the flow of access to two SM controllers 108 in the case of writing data in the two doubled shared memory sections 15 from SM access-control section 105. When doubling the two shared memory sections, one side is set as a master and another side is set as a slave. The data transfer control section 310 in SM access-control section 105 determines the royalty of the access path 2:137 to the SM controller 108 by the side of a master first by the Arbitration, and takes out the signal (REQ) which shows an access start by the control line 6; the control section 315 in the SM controller 108 by the side of a master (Step 1).

controller 108 by the side of a master. [0108] The data transfer control section 310 in SM access-control section 105 turns off the REQ signal to both SM controllers 108 a master and slave side, if the status is received from the SM controller 108 by the side of a slave (Steps 715a and 715b).  
[0109] When writing data in the doubled shared memory section 15 doubly, open both access paths 2:137 wide and there is nothing until it divides the shared memory section 15 into a master and a slave and a slave as mentioned above, it writes in data in order by the side of a master and a slave. The turn which writes in data by this and the writing of the data by the side of a slave finishes. The turn which writes in data written in doubly guaranteed, and while performing write-in processing of one data of the data written in doubly, another data can prevent being rewritten by the write request from other access paths.  
[0110] The data transfer control section 310 in SM access-control section 105 will report the end of access to the shared memory section 15 to a microprocessor 101 by the control line 111. The status is received.

[0116] In the disk array control unit 1 of the example 1 shown in [example 6] drawing 1, instead of connecting between SM access-control section 105 and the SM controller 108 by the access path 2:137, as shown in drawing 7, it connects through two share buses 130, two share buses 130 from SM access-control section 105 — it is alike, respectively and prepares a—one connection path at a time moreover — from the SM controller 108 — two share buses 130 is alike, respectively and prepares a—one connection path at a time SM access-control section 105 and the SM controller 108 have two arbiters respectively. Two arbiters are arbiters for the Arbitrations of two share buses, respectively. In the case of access, one of two or more arbiters from SM access-control section 105 to the SM controller 108 becomes a master, and it performs the Arbitration of the royalty of the share bus 130. And SM access-control section 105 which acquired the royalty accesses the SM controller 108. Moreover, it is satisfactory even if it carries out a direct file to the share bus 130 as an independent circuit, instead of preparing an — 105 and the SM controller 108.

arbiter in SM access-control section 103 and the shared bus connection by expanding the data width of face of a bus, and the access time to the shared memory section 15 can be shortened. [0118] Even if it carries out star connection (1 to 1 connection) of between the channel IF section 11 and the disk IF section 12, and the shared memory sections 15 like drawing 1 since the data width of face of the access path 2:137 in drawing 1 can be made smaller than the data width of face of the access path 2:137 in drawing 1.

width of face of an access path 0:135 more than double precision as mentioned above, there are few cases which the problem of the pin neck of LSI which mounts the shared memory section produces. But the problem that the number of an access path 2:137 increases too much, and an access path 2:137 cannot be mounted may arise. In such a case, the share bus connection of this example becomes effective.

[0119] In addition, at this example, although the selector section 13, the cache memory section 14, and the shared memory section 15 are doubled from a viewpoint of improvement in obstacle-proof nature, respectively, even if it does not double these, it cannot be overemphasized that an above-mentioned effect can be acquired.

[0120] [Effect of the invention] According to this invention, about the access path between channel IF section and disk IF section-cache memories, a throughput can be made high. Moreover, about the access path between channel IF section and disk IF section-shared memories, a throughput can be made high, and the access time can be shortened. By this, a throughput becomes possible [ offering the short disk array control unit of the response time highly ].

[Translation done.]

## \* NOTICES \*

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

## DESCRIPTION OF DRAWINGS

## [Brief Description of the Drawings]

[Drawing 1] Drawing showing the composition of the disk array control unit by this invention.  
 [Drawing 2] Drawing showing the composition of the conventional disk array control unit.  
 [Drawing 3] Drawing showing other composition of the conventional disk array control unit.  
 [Drawing 4] Drawing showing the composition of CM access-control section in the disk array control unit by this invention.  
 [Drawing 5] Drawing showing the composition of SM access-control section in the disk array control unit by this invention.  
 [Drawing 6] Drawing showing the composition of the selector section in the disk array control unit by this invention.  
 [Drawing 7] Drawing showing the composition of the cache memory section in the disk array control unit by this invention.  
 [Drawing 8] Drawing showing the composition of the shared memory section in the disk array control unit by this invention.  
 [Drawing 9] Drawing showing the procedure at the time of the writing of the data to the cache memory section.  
 [Drawing 10] Drawing showing the procedure at the time of the writing of the data to the shared memory section.  
 [Drawing 11] Drawing showing the procedure when writing data in the two doubled cache memory sections doubly.  
 [Drawing 12] Drawing showing the procedure when copying data to another cache memory section from the one cache memory section.  
 [Drawing 13] Drawing showing the procedure when writing data in the two doubled shared memory sections doubly.  
 [Drawing 14] Drawing showing the composition of mounting in the disk array control unit by this invention.  
 [Drawing 15] Drawing showing other composition in the disk array control unit by this invention.  
 [Drawing 16] Drawing showing other composition in the disk array control unit by this invention.  
 [Drawing 17] Drawing showing the composition of the disk array control unit by this invention.

## [Description of Notations]

1 — — a disk array control unit, the 11 — channel IF section, the 12 — disk IF section, and 13 — — the selector section, 14 — cache memory section, 15 — shared memory section, and 20 — — a magnetic disk unit, 50 — host computer, a 101 — microprocessor, and 102 — — Host IF, the 103 — drive IF, the 104 — CM access-control section, and 105 — — SM access-control section, the interior bus of 106 —, a 107 — CM controller, and 108 — —

[Translation done.]



## 【特許請求の範囲】

【請求項1】ホストコンピュータとのインターフェースを有する複数のチャネルインターフェース部と、磁気ディスク装置とのインターフェースを有する複数のディスクインターフェース部と、前記磁気ディスク装置に対しリード/ライトされるデータを一時的に格納するキャッシュメモリ部と、チャネルインターフェース部及びディスクインターフェース部と前記キャッシュメモリ部との間のデータ転送に関する制御情報を格納する共有メモリ部とを有し、各チャネルインターフェース部は、前記ホストコンピュータとのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行し、各ディスクインターフェース部は、前記磁気ディスク装置とのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行するディスクアレイ制御装置において、

前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間の接続形式が、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記共有メモリ部との間の接続形式と異なることを特徴とするディスクアレイ制御装置。

【請求項2】前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間はセレクタ部を介して接続され、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記共有メモリ部との間は、それぞれセレクタ部を介せず直接接続されていることを特徴とする請求項1に記載のディスクアレイ制御装置。

【請求項3】前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間はセレクタ部を介して接続され、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記共有メモリ部との間はそれぞれ1対1接続されていることを特徴とする請求項1に記載のディスクアレイ制御装置。

【請求項4】前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間はセレクタ部を介して接続され、前記複数のチャネルインターフェース部、前記複数のディスクインターフェース部、及び前記共有メモリ部は共有バスに接続されていることを特徴とする請求項1に記載のディスクアレイ制御装置。

【請求項5】前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間はスイッチを用いた相互結合網によって接続され、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記共有メモリ部との間はそれぞれ直接接続されていることを特徴とする請求項1に記載のディスクアレイ制御装置。

【請求項6】前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間はスイッチを用いた相互結合網によって接続され、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記共有メモリ部との間はそれぞれ1対1接続されていることを特徴とする請求項1に記載のディスクアレイ制御装置。

【請求項7】ホストコンピュータとのインターフェースを有する複数のチャネルインターフェース部と、磁気ディスク装置とのインターフェースを有する複数のディスクインターフェース部と、前記複数のチャネルインターフェース部と前記複数のディスクインターフェース部とに接続され、前記磁気ディスク装置に対しリード/ライトされるデータを一時的に格納するキャッシュメモリ部と、前記複数のチャネルインターフェース部と前記複数のディスクインターフェース部とに接続され、チャネルインターフェース部及びディスクインターフェース部と前記キャッシュメモリ部との間のデータ転送に関する制御情報を格納する共有メモリ部とを有する制御装置において、

前記キャッシュメモリ部に接続されるアクセスバスの本数は、前記共有メモリ部に接続されるアクセスバスの本数より少ないと特徴とするディスクアレイ制御装置。

【請求項8】セレクタ部をさらに有し、各チャネルインターフェース部及び各ディスクインターフェース部と前記セレクタ部とは、それぞれアクセスバスにより1対1に接続され、前記セレクタ部と前記キャッシュメモリ部とは、アクセスバスにより接続され、各チャネルインターフェース部及び各ディスクインターフェース部と前記セレクタ部とを接続するアクセスバスの総数は、前記セレクタ部と前記キャッシュメモリ部とを接続するアクセスバスの総数より多く、

各チャネルインターフェース部及び各ディスクインターフェース部と前記共有メモリ部との間はそれぞれアクセスバスにより1対1に接続されていることを特徴とする請求項7に記載のディスクアレイ制御装置。

【請求項9】ホストコンピュータとの複数のインターフェース部と、磁気ディスク装置との複数のインターフェース部と、磁気ディスク装置のデータを一時的に格納する物理的に独立したキャッシュメモリ部と、制御情報を格納する物理的に独立した共有メモリ部とを有し、前記ホストコンピュータとの各インターフェース部及び前記磁気ディスク装置との各インターフェース部は、それぞれマイクロプロセッサと、前記キャッシュメモリ部へのアクセスを制御するキャッシュメモリ制御部と、前記共有メモリ部へのアクセスを制御する共有メモリアクセス制御部とを有するディスクアレイ制御装置において、

各キャッシュメモリアクセス制御部と前記キャッシュメモリ部との間は、セレクタ部を介してアクセスバスによ

り接続されており、各共有メモリアクセス制御部と前記共有メモリ部との間は、それぞれアクセスバスにより直接接続されていることを特徴とするディスクアレイ制御装置。

【請求項10】ホストコンピュータとの複数のインターフェース部と、磁気ディスク装置との複数のインターフェース部と、磁気ディスク装置のデータを一時的に格納する物理的に独立したキャッシュメモリ部と、制御情報を格納する物理的に独立した共有メモリ部とを有し、前記ホストコンピュータとの各インターフェース部及び前記磁気ディスク装置との各インターフェース部は、それぞれマイクロプロセッサと、前記キャッシュメモリ部へのアクセスを制御するキャッシュメモリ制御部と、前記共有メモリ部へのアクセスを制御する共有メモリアクセス制御部とを有するディスクアレイ制御装置において、各キャッシュメモリアクセス制御部と前記キャッシュメモリ部との間は、スイッチを用いた相互結合網によって接続されており、各共有メモリアクセス制御部と前記共有メモリ部との間は、それぞれアクセスバスにより直接接続されていることを特徴とするディスクアレイ制御装置。

【請求項11】各キャッシュメモリ制御部と前記キャッシュメモリ部との間のアクセスバスの帯域幅を、前記ホストコンピュータと前記ホストコンピュータとのインターフェース部との間の最大の全帯域幅の2倍以上としたことを特徴とする請求項9または請求項10の何れかに記載のディスクアレイ制御装置。

【請求項12】前記共有メモリ部及び前記キャッシュメモリ部を複数有し、該複数の共有メモリ部及び前記複数のキャッシュメモリで二重化していることを特徴とする請求項1乃至請求項11の何れかに記載のディスクアレイ制御装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、データを複数の磁気ディスク装置に格納するディスクアレイ装置の制御装置に関する。

【0002】

【従来の技術】半導体記憶装置を記憶媒体とするコンピュータの主記憶のI/O性能に比べて、磁気ディスクを記憶媒体とするディスクサブシステム（以下「サブシステム」という。）のI/O性能は3～4桁程度小さく、従来からこの差を縮めること、すなわちサブシステムのI/O性能を向上させる努力がなされている。サブシステムのI/O性能を向上させるための1つの方法として、複数の磁気ディスク装置でサブシステムを構成し、データを複数の磁気ディスク装置に格納する、いわゆるディスクアレイと呼ばれるシステムが知られている。

【0003】図2は、従来のディスクアレイの構成を示す。ホストコンピュータ50とディスクアレイ制御装置

2との間のデータ転送を実行する複数のチャネルIF部11と、磁気ディスク装置20とディスクアレイ制御装置2間のデータ転送を実行する複数のディスクIF部12と、磁気ディスク装置20のデータを一時的に格納するキャッシュメモリ部14と、ディスクアレイ制御装置2に関する制御情報（例えば、チャネルIF部及びディスクIF部とキャッシュメモリ部14との間のデータ転送制御に関する情報）を格納する共有メモリ部15とを備え、キャッシュメモリ部14および共有メモリ部15は全てのチャネルIF部11及びディスクIF部12からアクセス可能な構成となっている。このディスクアレイでは、チャネルIF部11及びディスクIF部12と共有メモリ部15との間、及び、チャネルIF部11及びディスクIF部12とキャッシュメモリ部14との間は1対1に接続される。以下、このような接続形態をスター接続と呼ぶ。

【0004】チャネルIF部11は、ホストコンピュータ50と接続するためのインターフェース及びホストコンピュータ50に対する入出力を制御するマイクロプロセッサ（図示せず）を有している。また、ディスクIF部12は、磁気ディスク装置20と接続するためのインターフェース及び磁気ディスク装置20に対する入出力を制御するマイクロプロセッサ（図示せず）を有している。また、ディスクIF部12は、RAID機能の実行も行う。

【0005】図3は、他の従来のディスクアレイの構成を示す。ホストコンピュータ50とディスクアレイ制御装置3間のデータ転送を実行する複数のチャネルIF部11と、磁気ディスク装置20とディスクアレイ制御装置3間のデータ転送を実行する複数のディスクIF部12と、磁気ディスク装置20のデータを一時的に格納するキャッシュメモリ部14と、ディスクアレイ制御装置3に関する制御情報（例えば、チャネルIF部及びディスクIF部とキャッシュメモリ部14との間のデータ転送制御に関する情報）を格納する共有メモリ部15を備え、各チャネルIF部11及びディスクIF部12と共有メモリ部15間は共有バス130で接続され、各チャネルIF部11部及びディスクIF部12とキャッシュメモリ部14間は共有バス131で接続される。以下、このような接続形態を共有バス接続形式と呼ぶ。

【0006】

【発明が解決しようとする課題】ディスクアレイのアーキテクチャーをスケーラブルなものとするには、ディスク制御装置に接続するディスク容量（論理ボリューム数）に応じ、ディスクIF部を増設し、また、必要なホストコンピュータとのチャネル数に応じて、ディスクアレイ制御装置内のチャネルIF部を増設する必要がある。しかし、第3図に示した共有バス接続形式のディスクアレイ制御装置では、一旦実装した共有バスの転送能力をチャネルIF部、ディスクIF部の増設に応じて変

更することはできないので、チャネルIF部、ディスクIF部の増設に柔軟に対応することが困難である。

【0007】また、第3図に示した共有バス接続形式のディスクアレイ制御装置では、ホストコンピュータとディスクアレイ制御装置との間のデータ転送を実行するチャネルIF部に設けられたホストコンピュータに対する入出力を制御するマイクロプロセッサ、及び磁気ディスク装置とディスクアレイ制御装置との間のデータ転送を実行するディスクIF部に設けられた磁気ディスク装置に対する入出力を制御するマイクロプロセッサに高性能なプロセッサを使用した場合に、これらのプロセッサの性能に比べて、共有バスの転送能力がボトルネックになり、プロセッサの高速化に追従することが困難となる。

【0008】さらに、第3図に示した共有バス接続形式では、共有バスに接続された複数のチャネルIF部（または複数のディスクIF部）の何れかのチャネルIF部（またはディスクIF部）に障害が発生した場合に、障害の発生したチャネルIF部（またはディスクIF部）を特定することが困難である。

【0009】一方、第2図に示したスター接続形式のディスクアレイ制御装置では、共有メモリ部またはキャッシュメモリ部に接続したアクセスバス数に比例して内部バス性能が増加させることができるので、チャネルIF部、ディスクIF部の増設、または使用するプロセッサの性能に応じて、内部バス性能を増加させることができるので、チャネルIF部及びディスクIF部とキャッシュメモリ部との間、チャネルIF部及びディスクIF部と共有メモリ部との間がスター接続されているため、障害の発生したチャネルIF部（またはディスクIF部）を特定することも容易である。

【0010】スター接続形式のディスクアレイ制御装置では、搭載されるチャネルIF部またはディスクIF部の数を増やした場合、チャネルIF部及びディスクIF部とキャッシュメモリ部との間、及びチャネルIF部及びディスクIF部と共有メモリとの間のアクセスバス数も増えることになる。また、ホストコンピュータとディスクアレイ制御装置との間の接続にファイバチャネル等の高速チャネルの採用等により、ディスクアレイ制御装置に要求されるスループットはさらに増大する方向にあり、このスループットの向上の要求を満たすためには、チャネルIF部及びディスクIF部とキャッシュメモリ部との間、及びチャネルIF部及びディスクIF部と共有メモリとの間のアクセスバス数を増やし、内部バス性能を向上させることが必要となる。

【0011】しかし、キャッシュメモリに格納される1つのデータのデータ量は、共有メモリに格納される1つの制御情報のデータ量よりもかなり大きい。一例を挙げれば、メインフレームに接続されるディスク制御装置では、キャッシュメモリに格納される1つのデータは数Kバイト程度（例えば2Kバイト）であるのに対し、共有

メモリに格納される1つの制御情報は数バイト程度（例えば4バイト）である。また、オープン系のホストコンピュータに接続されるディスク制御装置では、キャッシュメモリに格納される1つのデータは数十バイト程度（例えば64バイト）であるのに対し、共有メモリに格納される1つの制御情報は数バイト程度（例えば4バイト）である。したがって、チャネルIF部及びディスクIF部とキャッシュメモリ部との間で転送されるデータ量は、チャネルIF部及びディスクIF部と共有メモリ部との間で転送されるデータ量に比べ、かなり多いので、チャネルIF部及びディスクIF部とキャッシュメモリ部との間のアクセスバスのデータ幅は、チャネルIF部及びディスクIF部と共有メモリ部との間のアクセスバスのデータ幅より広くとる必要がある。例えば、前者のアクセスバスは、16ビット幅のバスで構成され、後者は、4ビット幅のバスで構成される。そのため、チャネルIF部及びディスクIF部とキャッシュメモリ部との間のアクセスバスの本数を増やすと、それらのアクセスバスを接続するキャッシュメモリ部のLSIのピン数が不足するという問題が生じる。また、ディスクアレイ制御装置のホストコンピュータへの応答時間を短くするためには、共有メモリ部に格納された制御情報へのアクセス時間をできるだけ短くすることも必要である。

【0012】そこで、本発明の目的は、キャッシュメモリ及び共有メモリに格納されるデータの特性及びこれらのメモリへのアクセス特性を考慮した、スループットの高く、ディスクアレイ制御装置、及びそれを用いたサブシステムを提供することにある。

【0013】より具体的には、本発明の目的は、チャネルIF部及びディスクIF部とキャッシュメモリ部との間のアクセスバスはスループットが高く、チャネルIF部及びディスクIF部と共有メモリとの間のアクセスバスはスループットが高く、かつアクセス時間が短いディスクアレイ制御装置、及びそれを用いたサブシステムを提供することにある。

【0014】

【課題を解決するための手段】上記目的は、ホストコンピュータとのインターフェースを有する複数のチャネルインターフェース部と、磁気ディスク装置とのインターフェースを有する複数のディスクインターフェース部と、前記磁気ディスク装置に対しリード/ライトされるデータを一時的に格納するキャッシュメモリ部と、チャネルインターフェース部及びディスクインターフェース部と前記キャッシュメモリ部との間のデータ転送に関する制御情報を格納する共有メモリ部とを有し、各チャネルインターフェース部は、前記ホストコンピュータとのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行し、各ディスクインターフェース部は、前記磁気ディスク装置とのインターフェースと前記キャッシュメモリ部との間のデータ転送を実行するディスク

アレイ制御装置において、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間の接続形式が、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記共有メモリ部との間の接続形式と異なることを特徴とするディスクアレイ制御装置により達成される。

【0015】好ましくは、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間はセレクタ部を介して接続し、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記共有メモリ部との間は、それぞれセレクタ部を介せず直接接続する。

【0016】また、好ましくは、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間はセレクタ部を介して接続し、前記複数のチャネルインターフェース部、及び前記共有メモリ部は共有バスに接続する。

【0017】また、好ましくは、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記キャッシュメモリ部との間はスイッチを用いた相互結合網によって接続し、前記複数のチャネルインターフェース部及び前記複数のディスクインターフェース部と前記共有メモリ部との間はそれぞれ直接接続する。

【0018】その他、本願が開示する課題、及びその解決方法は、発明の実施形態の欄及び図面により明らかにされる。

【0019】

【発明の実施の形態】以下、本発明の実施例を図面を用いて説明する。

【0020】【実施例1】図1に、本発明の一実施例を示す。

【0021】ディスクアレイ制御装置1は、ホストコンピュータ50との2つのインターフェース部（チャネルIF部）11と、磁気ディスク装置20との2つのインターフェース部（ディスクIF部）12と、2つのセレクタ部13と、2つのキャッシュメモリ部14と、2つの共有メモリ部15と、アクセスバス0:135と、アクセスバス1:136と、アクセスバス2:137とを有する。

【0022】チャネルIF部11は、ホストコンピュータ50との2つのIF（ホストIF）102と、ホストコンピュータ50に対する入出力を制御する2つのマイクロプロセッサ101と、キャッシュメモリ部14へのアクセスを制御するアクセス制御部（CMアクセス制御部）104と、共有メモリ部15へのアクセスを制御するアクセス制御部（SMアクセス制御部）105とを有する。

し、ホストコンピュータ50とキャッシュメモリ部14間のデータ転送、及びマイクロプロセッサ101と共有メモリ部15間の制御情報の転送を実行する。マイクロプロセッサ101及びホストIF102は内部バス106によって接続され、CMアクセス制御部104は2つのホストIF102に直接接続されている。また、SMアクセス制御部105は2つのマイクロプロセッサ101に直接接続されている。

【0023】ディスクIF部12は、磁気ディスク装置20との2つのIF（ドライブIF）103と、磁気ディスク装置20に対する入出力を制御する2つのマイクロプロセッサ101と、キャッシュメモリ部14への1つのアクセス制御部（CMアクセス制御部）104と、共有メモリ部15への1つのアクセス制御部（SMアクセス制御部）105を有し、磁気ディスク装置20とキャッシュメモリ部14間のデータ転送、及びマイクロプロセッサ101と共有メモリ部15間の制御情報の転送を実行する。マイクロプロセッサ101及びドライブIF103は内部バス106によって接続され、CMアクセス制御部104は2つのドライブIF103に直接接続されている。また、SMアクセス制御部105は2つのマイクロプロセッサ101に直接接続されている。ディスクIF部はRAID機能の実行も行う。

【0024】キャッシュメモリ部14は、キャッシュメモリ（CM）コントローラ107とメモリモジュール109を有し、磁気ディスク装置20へ記録するデータを一時的に格納する。

【0025】共有メモリ部15は、共有メモリ（SM）コントローラ108とメモリモジュール109とを有し、ディスクアレイ制御装置1の制御情報（例えば、チャネルIF部及びディスクIF部とキャッシュメモリ部14との間のデータ転送制御に関する情報）等を格納する。

【0026】CMアクセス制御部104には2本のアクセスバス0:135を接続し、それらを2つの異なるセレクタ部13にそれぞれ接続する。セレクタ部13には2本のアクセスバス1:136を接続し、それらを2つの異なるCMコントローラ107にそれぞれ接続する。したがってCMコントローラ107には、2つのセレクタ部から1本ずつ、計2本のアクセスバス1:136が接続される。こうすることにより、1つのCMアクセス制御部104から1つのCMコントローラ107へのアクセスルートが2つとなる。これにより、1つのアクセスバスまたはセレクタ部13に障害が発生した場合でも、もう1つのアクセスルートによりキャッシュメモリ部14へアクセスすることが可能となるため、耐障害性を向上させることができる。

【0027】SMアクセス制御部105には2本のアクセスバス2:137を接続し、そのアクセスバス2:137を2つの異なるSMコントローラにそれぞれ接続す

る。したがって、SMコントローラ108には、2つのチャネルIF部11及び2つのディスクIF部12から1本ずつ、計4本のアクセスバス2:137が接続される。本実施例では、1つのSMアクセス制御部105と1つのSMコントローラ108の間には1本のアクセスバス2:137を接続したが、このアクセスバス2:137を2本に増やすことにより、1つのSMアクセス制御部105から1つのSMコントローラ108へのアクセスルートが2つとなるため、耐障害性を向上させることができる。

【0028】セレクタ部13には、2つのチャネルIF部11と、2つのディスクIF部12からそれぞれ1本ずつ、計4本のアクセスバス0:135が接続される。また、セレクタ部13には、2つのキャッシュメモリ部14へのアクセスバス1:136が1本ずつ、計2本接続される。

【0029】アクセスバス0:135とアクセスバス1:136の間に上記のようなバス数の関係があるため、セレクタ部13ではチャネルIF部11及びディスクIF部12からの4本のアクセスバス0:135からの要求の内、キャッシュメモリ部14へのアクセスバス1:136の数に相当する2個だけを選択して実行する機能を持つ。

【0030】本実施例の大きな特徴は、チャネルIF部11及びディスクIF部12とキャッシュメモリ部14との間の接続形式と、チャネルIF部11及びディスクIF部12と共有メモリ部15との間の接続形式とが異なる点にある。このような構成にした理由を図1及び図2を用いて説明する。図2に示したスター接続形式のディスクアレイ制御装置2では、ディスクアレイ制御装置2に搭載されるチャネルIF部11またはディスクIF部12の数を増やした場合、チャネルIF部11及びディスクIF部12とキャッシュメモリ部14との間、及びチャネルIF部11及びディスクIF部12と共有メモリ部15との間のアクセスバス数も増えることになる。また、スター接続形式のディスクアレイ制御装置2において、スループットを向上させるためには、チャネルIF部11及びディスクIF部12とキャッシュメモリ部14との間、及びチャネルIF部11及びディスクIF部12と共有メモリ部15との間のアクセスバス数を増やし、内部バス性能を向上させることができある。

【0031】しかし、キャッシュメモリ部14に格納される1つのデータのデータ量は、共有メモリ部15に格納される1つの制御情報のデータ量よりもかなり大きい。一例を挙げれば、メインフレームに接続されるディスク制御装置では、キャッシュメモリ部14に格納される1つのデータは数Kバイト程度（例えば2Kバイト）であるのに対し、共有メモリ部15に格納される1つの制御情報は数バイト程度（例えば4バイト）である。ま

た、オープン系のホストコンピュータに接続されるディスク制御装置では、キャッシュメモリ部14に格納される1つのデータは数十バイト程度（例えば64バイト）であるのに対し、共有メモリ部15に格納される1つの制御情報は数バイト程度（例えば4バイト）である。したがって、チャネルIF部11及びディスクIF部12とキャッシュメモリ部14との間で転送されるデータ量は、チャネルIF部11及びディスクIF部12と共有メモリ部15との間で転送されるデータ量に比べかなり多いので、チャネルIF部11及びディスクIF部12とキャッシュメモリ部14との間のアクセスバスのデータ幅は、チャネルIF部11及びディスクIF部12と共有メモリ部15との間のアクセスバスのデータ幅より広くとる必要がある。例えば、前者のアクセスバスは16ビット幅のバスで構成され、後者は、8ビット幅のバスで構成される。そのため、チャネルIF部11及びディスクIF部12とキャッシュメモリ部14との間のアクセスバス本数を増やすと、キャッシュメモリ部14内のキャッシュメモリコントローラ（図2ではキャッシュメモリコントローラを図示していない）のLSIのピン数不足、またはキャッシュメモリ部14を実装するパッケージにおいてコネクタのピン数不足という問題が生じる。そこで、本実施例では、図1に示すように、チャネルIF部11及びディスクIF部12とキャッシュメモリ部14との間をセレクタ部13を介して接続することにより、キャッシュメモリ部14に直接接続されるアクセスバス数を削減している。

【0032】一方、上述したように、共有メモリ部13へ格納する1つの制御情報のデータ長はキャッシュメモリ部14に格納する1つのデータのデータ長に比べかなり小さいので、チャネルIF部11及びディスクIF部12と共有メモリ部15との間のアクセスバスのデータ幅は、チャネルIF部11及びディスクIF部12とキャッシュメモリ部14との間のアクセスバスのデータ幅の半分以下とすることが可能である。したがって、共有メモリ部15へのアクセスバス数を増やしても共有メモリ部内の共有メモリメモリコントローラ（図2では共有メモリコントローラを図示していない）のLSIのピン数不足等の問題が生じることは少ない。

【0033】また、ディスクアレイ制御装置1のホストコンピュータ50への応答時間を短くするためにには、共有メモリ部15に格納される制御情報へのアクセス時間をできるだけ短くする必要もある。しかし、図1に示したCMアクセス制御部104とCMコントローラ107間のように、SMアクセス制御部105とSMコントローラ108との間をセレクタ部を介して接続すると、セレクタ部での処理のオーバーヘッドにより、共有メモリ部15に格納される制御情報へのアクセス時間を短くすることができない。

【0034】そこで、本実施例では、チャネルIF部1

1及びディスクIF部12と共有メモリ部15との間をセレクタ部を介さず直接接続することにより、チャネルIF部11及びディスクIF部12と共有メモリ部15間に複数のアクセスパスを設け、スループットを向上させる一方、セレクタ部での処理オーバーヘッドをなくし、チャネルIF部11、及びディスクIF部12から共有メモリ部15へのアクセス時間を短縮している。

【0035】なお、本実施例では、耐障害性の向上という観点から、セレクタ部13、キャッシュメモリ部14、及び共有メモリ部15をそれぞれ二重化しているが、これらを二重化しなくても、上述の効果を得られるることは言うまでもない。

【0036】図4は、CMアクセス制御部104内の構成を示している。CMアクセス制御部104は、セレクタ302と、アドレス、コマンド、データを一時格納するパケットバッファ303と、セレクタ部13に繋がるアクセスパス0:135とのバスIF301と、データのエラーチェック部300と、データ転送制御部310を有する。セレクタ302の2つのポートはデータ線210でホストIF102あるいはドライブIF103に接続される。また、セレクタ302の他の2つのポートはバスIF301に接続される。バスIF301はアクセスパス0:135でセレクタ部13に接続される。データ転送制御部310は、制御線1:211でホストIF102あるいはドライブIF103に接続され、制御線2:212でセレクタ部13内のデータ転送制御部315に接続される。また、データ転送制御部310は、アービタ308によりホストIF102あるいはドライブIF103からのアクセス要求のアービトレーションを行い、セレクタ302の切り替えを行う。

【0037】図6は、セレクタ部13内の構成を示している。セレクタ部13は、チャネルIF部11及びディスクIF部12に繋がるアクセスパス0:135との4つのバスIF301と、CMコントローラ107に繋がるアクセスパス1:136との2つのバスIF301と、両者間を互いに接続するセレクタ306と、パケットバッファ303と、データのエラーチェック部300と、CMアクセス制御部104から送出されたアドレス及びコマンドを解析するアドレス・コマンド(adr、cmd)解析部305と、データ転送制御部315を有する。データ転送制御部315は、制御線2:212でCMアクセス制御部104内のデータ転送制御部310に接続され、制御線3:213でCMコントローラ107内のデータ転送制御部315に接続される。また、データ転送制御部315は、アービタ308により、adr、cmd解析部305で解析した4つのアクセスパス0:135からのアクセス要求のアービトレーションを行い、セレクタ306の切り替えを行う。パケットバッファ303は、アクセスパス0:135側のバスとアクセスパス1:136側のバスでデータ転送速度に差がある

る場合、速度差を吸収するために、転送するデータの一部または全部をバッファリングする。

【0038】adr、cmd解析部305は、アドレス及びコマンドを格納するバッファと、adr抽出部と、cmd抽出部を有する(図示していない)。adr、cmd解析部305では、CMアクセス制御部104に接続される4本のアクセスパス0:135それぞれに1つずつ割り当てられたバッファに、アドレス、コマンドを格納する。adr抽出部及びcmd抽出部では、アクセスするCMコントローラ107とアクセスの種類を割り出し、データ転送制御部315内のアービタ308へ送出する。

【0039】図7は、キャッシュメモリ部14内の構成を示している。キャッシュメモリ部14は、CMコントローラ107とメモリモジュール109を有する。CMコントローラ107は、セレクタ部13に繋がるアクセスパス1:136との2つのバスIF301と、セレクタ304と、データを一時格納するパケットバッファ303と、データのエラーチェック部300と、メモリモジュール109へのアクセスを制御するメモリ制御部307と、CMアクセス制御部104から送出されたアドレス及びコマンドを解析するadr、cmd解析部305と、データ転送制御部315を有する。データ転送制御部315は、制御線3:213でセレクタ部13内のデータ転送制御部315に接続される。また、データ転送制御部315は、アービタ308により、adr、cmd解析部305で解析した2本のアクセスパス1:136からのアクセス要求のアービトレーションを行い、セレクタ304の切り替えを行う。

【0040】adr、cmd解析部305は、バッファと、adr抽出部と、cmd抽出部を有する(図示していない)。adr、cmd解析部305では、CMコントローラ107に接続される2本のアクセスパス1:136それぞれに1つずつ割り当てられたバッファに、アドレス、コマンドを格納する。adr抽出部及びcmd抽出部では、アクセスするメモリのアドレスとアクセスの種類を割り出し、メモリ制御部307へ送出する。また、2本のアクセスパス1:136からのアクセス要求をデータ転送制御部315内のアービタ308へ送出する。

【0041】次に、キャッシュメモリ部14へのアクセス時の手順について述べる。キャッシュメモリ部14へアクセスする場合、マイクロプロセッサ101は、ホストIF102あるいはドライブIF103へ、キャッシュメモリ部14へのアクセス開始を指示する。

【0042】アクセス開始の指示を受けたホストIF102あるいはドライブIF103は、制御線1:211によりCMアクセス制御部104内のデータ転送制御部310へアクセス開始を示す信号を送出する。それとともに、データ線210を通してアドレス、コマンド、デ

ータ（データの書き込み時のみ）を送出する。

【0043】CMアクセス制御部104は、データ線210を通して送られてきたアドレス、コマンド、データ（データの書き込み時のみ）をパケットバッファ303に格納する。データ転送制御部310はアービトレーションを行ってバスIF301の使用権を決定し、セレクタ302を切り替える。

【0044】図9は、キャッシュメモリ部14へデータを書き込む場合の、CMアクセス制御部104からCMコントローラ107へのアクセスの流れを示している。CMアクセス制御部104内のデータ転送制御部310は、アービトレーションによってアクセスバス0:135の使用権が決定されると、制御線2:212によってセレクタ部13内のデータ転送制御部315へアクセス開始を示す信号（REQ）を出す（ステップ501）。続いて、アドレス及びコマンドを送出する（ステップ502）。

【0045】セレクタ部13内のデータ転送制御部315は、CMアクセス制御部104からREQ信号を受け取ると、次にアクセスバス0:135を通して送られてくるアドレス及びコマンドを受信し、addr、cmd解析部305で解析したアクセス要求に基づいてアービトレーションを行う（ステップ503）。アービトレーションの結果、アクセスバス1:136への接続権を得たら、データ転送制御部315はセレクタ306を切り替える（ステップ504）とともに、制御線2:212により、CMアクセス制御部104内のデータ転送制御部310へ、アクセスバス1:136への接続権が得られたことを示す信号（ACK）を返す（ステップ505）。次にデータ転送制御部315は、制御線3:213によってCMコントローラ107内のデータ転送制御部315へアクセス開始を示す信号（REQ）を出す（ステップ506）。続いて、アドレス及びコマンドを送出する（ステップ507）。

【0046】CMアクセス制御部104はACK信号を受けると、パケットバッファ303からデータを読み出し、セレクタ302、バスIF301を介してアクセスバス0:135へ送出する。セレクタ部13は、アクセスバス0:135を通して送られてきたデータを、バスIF301及びセレクタ306を介してアクセスバス1:136へ送出する（ステップ509）。

【0047】CMコントローラ107内のデータ転送制御部315は、制御線3:213によってREQ信号を受け取ると、次にアクセスバス1:136を通して送られてくるアドレス及びコマンドを受信し、addr、cmd解析部305で解析したアクセス要求に基づいてアービトレーションを行い（ステップ508）、セレクタ304を切り替える。アクセスバス1:136を通して送られてくるデータはパケットバッファ303に格納する。アービトレーションの結果、メモリモジュール109へのアクセス権を得たら、メモリの制御情報をメモリ制御部307へ送出し、メモリアクセスのための前処理を行う（ステップ510）。次に、パケットバッファ303からデータを読み出し、セレクタ304を介してメモリモジュール109へ書き込む（ステップ511）。

【0048】メモリモジュール109へのアクセスが終了すると、メモリアクセスの後処理を行い、データ転送制御部315においてアクセス状況を示すステータス（STATUS）を生成する（ステップ512）。次に、ステータスをセレクタ部13を介してCMアクセス制御部104へ送出する（ステップ513）。セレクタ部13内のデータ転送制御部315はステータスを受け取ると、CMコントローラ107へのREQ信号をオフする（ステップ514）。CMアクセス制御部104内のデータ転送制御部310はステータスを受け取ると、セレクタ部13へのREQ信号をオフする（ステップ515）。セレクタ部13内のデータ転送制御部315はCMアクセス制御部104からのREQ信号のオフを確認すると、CMアクセス制御部104へのACK信号をオフする（ステップ516）。

【0049】CMアクセス制御部104内のデータ転送制御部310はステータスを受け取ると、制御線1:211により、ホストIF102あるいはドライブIF103へキャッシュメモリ部14へのアクセスの終了を報告する。

【0050】キャッシュメモリ部14からデータを読み出す場合の、CMアクセス制御部104からCMコントローラ107へのアクセスの流れは、ステップ501から508までとステップ512以降は、データの書き込みの場合と同じである。

【0051】ここでCMアクセス制御部104は、ステップ505でACK信号を受けると、データの受信待ち状態に入る。

【0052】ステップ508でメモリアクセス権を得ると、CMコントローラ107はメモリモジュール109からデータを読み出し、セレクタ304、バスIF301を介してアクセスバス1:136にデータを送出する。

【0053】セレクタ部13は、アクセスバス1:136を通してデータを受信すると、バスIF301及びセレクタ306を介してアクセスバス0:135にデータを送出する。

【0054】CMアクセス制御部104は、アクセスバス0:135を通してデータを受信すると、セレクタ302、データ線210を介してホストIF102あるいはドライブIF103へデータを送出する。

【0055】図5は、SMアクセス制御部105内の構成を示している。SMアクセス制御部104は、セレクタ302と、アドレス、コマンド、データを一時格納するパケットバッファ303と、SMコントローラ108

に繋がるアクセスバス2:137とのバスIF301と、データのエラーチェック部300と、データ転送制御部310を有する。セレクタ302の2つのポートはデータ線220でマイクロプロセッサ101に接続される。また、セレクタ302の他の2つのポートはバスIF301に接続される。バスIF301はアクセスバス2:137でSMコントローラ108に接続される。データ転送制御部310は、制御線5:221でマイクロプロセッサ101に接続され、制御線6:222でSMコントローラ108内のデータ転送制御部315に接続される。また、データ転送制御部310は、アビタ308によりマイクロプロセッサ101からのアクセス要求のアビトレーションを行い、セレクタ302の切り替えを行う。

【0056】図8は、共有メモリ部15内の構成を示している。共有メモリ部15は、SMコントローラ108とメモリモジュール109を有する。SMコントローラ108は、SMアクセス制御部105に繋がるアクセスバス2:137との4つのバスIF301と、セレクタ309と、データを一時格納するパケットバッファ303と、データのエラーチェック部300と、メモリモジュール109へのアクセスを制御するメモリ制御部307と、SMアクセス制御部105から送出されたアドレス及びコマンドを解析するadr、cmd解析部305と、データ転送制御部315を有する。データ転送制御部315は、制御線6:222でSMアクセス制御部105内のデータ転送制御部310に接続される。また、データ転送制御部315は、アビタ308により、adr、cmd解析部305で解析した4本のアクセスバス2:137からのアクセス要求のアビトレーションを行い、セレクタ309の切り替えを行う。

【0057】adr、cmd解析部305は、バッファと、adr抽出部と、cmd抽出部を有する(図示していない)。adr、cmd解析部305では、SMコントローラ108に接続される4本のアクセスバス2:137それぞれに1つずつ割り当てられたバッファに、アドレス、コマンドを格納する。adr抽出部及びcmd抽出部では、アクセスするメモリのアドレスとアクセスの種類を割り出し、メモリ制御部307へ送出する。また、4本のアクセスバス2:137からのアクセス要求をデータ転送制御部315内のアビタ308へ送出する。

【0058】次に、共有メモリ部15へのアクセス時の手順について述べる。共有メモリ部15へアクセスする場合、マイクロプロセッサ101は、制御線5:221によりSMアクセス制御部105内のデータ転送制御部310へアクセス開始を示す信号を送出する。それとともに、データ線220を通してアドレス、コマンド、データ(データの書き込み時のみ)を送出する。

【0059】SMアクセス制御部105は、データ線2

20を通して送られてきたアドレス、コマンド、データ(データの書き込み時のみ)をパケットバッファ303に格納する。データ転送制御部310はアビトレーションを行ってバスIF301の使用権を決定し、セレクタ302を切り替える。

【0060】図10は、共有メモリ部15へデータを書き込む場合の、SMアクセス制御部105からSMコントローラ108へのアクセスの流れを示している。SMアクセス制御部105内のデータ転送制御部310は、アビトレーションによってアクセスバス2:137の使用権が決定されると、制御線6:222によってSMコントローラ108へアクセス開始を示す信号(REQ)を出す(ステップ601)。続いて、アドレス、コマンド、及びデータを連続して送出する(ステップ602)。

【0061】SMコントローラ108内のデータ転送制御部315は、制御線6:222によってREQ信号を受け取ると、次にアクセスバス2:137を通して送られてくるアドレス、コマンド、及びデータを受信する。アドレスとコマンドは、adr、cmd解析部305で解析し、アクセス要求に基づいてアビトレーションを行い(ステップ603)、セレクタ309を切り替える。データはパケットバッファ303に格納する。アビトレーションの結果、メモリモジュール109へのアクセス権を得たら、メモリの制御情報をメモリ制御部307へ送出し、メモリアクセスのための前処理を行う(ステップ604)。次に、パケットバッファ303からデータを読み出し、セレクタ309を介してメモリモジュール109へ書き込む(ステップ605)。

【0062】メモリモジュール109へのアクセスが終了すると、メモリアクセスの後処理を行い、データ転送制御部315においてアクセス状況を示すステータス(STATUS)を生成する(ステップ606)。次に、ステータスをSMアクセス制御部105へ送出する(ステップ607)。SMアクセス制御部105内のデータ転送制御部310はステータスを受け取ると、SMコントローラ108へのREQ信号をオフする(ステップ608)。

【0063】SMアクセス制御部105内のデータ転送制御部310はステータスを受け取ると、制御線5:221により、マイクロプロセッサ101へ共有メモリ部15へのアクセスの終了を報告する。

【0064】共有メモリ部15からデータを読み出す場合のSMアクセス制御部105からSMコントローラ108へのアクセスの流れは、ステップ601から604までとステップ606以降は、データの書き込みの場合と同じである。

【0065】ステップ604でメモリアクセスの前処理を行った後、SMコントローラ108はメモリモジュール109からデータを読み出し、セレクタ309、バス

I F 3 0 1 を介してアクセスバス 2 : 1 3 7 にデータを送出する。

【0066】SMアクセス制御部105は、アクセスバス2:137を通してデータを受信すると、セレクタ302、データ線220を介してマイクロプロセッサ101へデータを送出する。

【0067】本実施例のディスクアレイ制御装置1では、ホストコンピュータ50とのチャネルを2つ有するチャネルI F部12を複数搭載しており、それらのチャネルをそれぞれ異なるホストコンピュータに接続することが可能である。そうした場合、ホストコンピュータに接続した各チャネルからの要求全てを並列に処理する必要がある。

【0068】ところでディスクアレイ制御装置1では、ホストコンピュータ50へデータを読み出す場合、磁気ディスク装置20に格納されたデータをディスクI F部12を介してキャッシュメモリ部14へ書き込み、そのデータをキャッシュメモリ部14から読み出して、チャネルI F部11を介してホストコンピュータ50へ送る。またホストコンピュータ50からディスクアレイ制御装置1へデータを書き込む場合は、ホストコンピュータ50からチャネルI F部12へ送られてきたデータをキャッシュメモリ部14へ書き込み、そのデータをキャッシュメモリ部14から読み出して、ディスクI F部12を介して磁気ディスク装置20に書き込む。さらにデータのパリティを生成して磁気ディスク装置20に書き込むため、ディスクI F部12とキャッシュメモリ部14間でさらに2~3回のアクセスが行われる。

【0069】したがって、ホストコンピュータに接続した各チャネルからの要求全てを並列に処理するためには、チャネルI F部11及びディスクI F部12とキャッシュメモリ部14間のスループットをホストコンピュータ50とチャネルI F部11間の最大のスループットの2倍以上にしなければならない。

【0070】本実施例では、CMアクセス制御部104とCMコントローラ107間のアクセスバスの帯域幅、及びCMコントローラ107とメモリモジュール109間の全帯域幅を、チャネルI F部11とホストコンピュータ50間の最大の帯域幅の2倍以上に設定する。これにより、全チャネルI F部11を並列に動作させることが可能となる。

【0071】本実施例によれば、チャネルI F部11及びディスクI F部12とキャッシュメモリ部14間のスループットを高くすること、チャネルI F部11及びディスクI F部12と共有メモリ部15間のスループットを高く、且つアクセス時間を短くすることの両方が可能となる。これによって、スループットが高く、且つ応答時間の短いディスクアレイ制御装置を提供できる。

【0072】ここで、図15に示すように、チャネルI F部11及びディスクI F部12とキャッシュメモリ部

14間をスイッチ(SW)16を用いた相互結合網140で接続する。この場合にも、図1に示したセレクタ部13を介して接続した構成と同様に、キャッシュメモリ部14へ複数のアクセスバスを設けることができるため、スループットを高めることができる。

【0073】また図16に示すように、1つのCMアクセス制御部104へ接続されるアクセスバス0:135の本数を図1の構成の倍の4本に増やしたディスクアレイ制御装置1においても、本実施例を実施する上で問題はない。ホストI F及びドライブI Fとして、今後はファイバーチャネル等のスループットが100MB/s以上の高速I Fが使用されることが多くなると考えられる。チャネルI F部11及びディスクI F部12でのスループットのバランスを考えると、1つのCMアクセス制御部104に繋がる全アクセスバス0:135のスループットは、チャネルI F部11内の全ホストI F102、またはディスクI F部12内の全ドライブI F103のスループットと同等以上にする必要がある。上記のようにファイバーチャネル等の高速I Fを使用する場合は、図16に示すように、1つのCMアクセス制御部104に繋がるアクセスバス0:135の本数を増やすことで、アクセスバス0:135のスループットをホストI F102あるいはドライブI F103のスループット以上にすることができる。

【0074】なお、図15及び図16のディスクアレイ制御装置では、耐障害性の向上という観点から、セレクタ部13、キャッシュメモリ部14、及び共有メモリ部15をそれぞれ二重化しているが、これらを二重化しなくても、上述の効果を得られることは言うまでもない。

【0075】【実施例2】図1において、2つのキャッシュメモリ部14間で、メモリ領域の全部または、一部を二重化し、キャッシュメモリ部14へのデータの書き込み時に二重化した2つの領域に同じデータを書き込むことにより、データの信頼性を上げることが可能となる。

【0076】二重化した2つのキャッシュメモリ部14へデータを書き込む場合の手順は以下のようになる。

【0077】アクセス開始の指示を受けたホストI F102あるいはドライブI F103は、制御線1:211によりCMアクセス制御部104内のデータ転送制御部310へアクセス開始を示す信号を送出する。それとともに、データ線210を通して2つのアドレス、コマンドと、1つのデータを送出する。

【0078】CMアクセス制御部104は、データ線210を通して送られてきた2つのアドレス、コマンドと、1つのデータをパケットバッファ303に格納する。データ転送制御部310はアビトリレーションを行ってバスI F301の使用権を決定し、セレクタ302を切り替える。

【0079】図11は、二重化した2つのキャッシュメ

モリ部14へデータを書き込む場合の、CMアクセス制御部104から2つのCMコントローラa、b:107へのアクセスの流れを示している。CMアクセス制御部104内のデータ転送制御部310は、アービトレーションによってアクセスパス0:135の使用権が決定されると、制御線2:212によってセレクタ部13内のデータ転送制御部315へアクセス開始を示す信号(REQ)を出す(ステップ701)。続いて、アドレス及びコマンドを2つ連続して送出する(ステップ702)。

【0080】セレクタ部13内のデータ転送制御部315は、CMアクセス制御部104からREQ信号を受けると、次にアクセスパス0:135を通して送られてくるアドレス及びコマンドを受信し、adr、cmd解析部305で解析したアクセス要求に基づいてアービトレーションを行う(ステップ703)。アービトレーションの結果、CMコントローラa、b:107への2本のアクセスパス1:136への接続権の両方を得たら、データ転送制御部315はセレクタ306を切り替える(ステップ704)とともに、制御線2:212により、CMアクセス制御部104内のデータ転送制御部310へ、CMコントローラa、b:107への接続権が得られたことを示す信号(ACK)を返す(ステップ705)。次にデータ転送制御部315は、制御線3:213によって2つのCMコントローラ107内のデータ転送制御部315へアクセス開始を示す信号(REQ)を出す(ステップ706a、706b)。続いて、CMコントローラa、b:107へアドレス及びコマンドを1つずつ送出する(ステップ707a、707b)。

【0081】CMアクセス制御部104はACK信号を受けると、パケットバッファ303からデータを読み出し、セレクタ302、バスIF301を介してアクセスパス0:135へ送出する(ステップ709)。セレクタ部13は、アクセスパス0:135を通して送られてきた1つのデータを、バスIF301及びセレクタ306を介して2つのアクセスパス1:136の両方へ送出する(ステップ709a、709b)。

【0082】CMコントローラa、b:107内のデータ転送制御部315は、制御線3:213によってREQ信号を受け取ると、次にアクセスパス1:136を通して送られてくるアドレス及びコマンドを受信し、adr、cmd解析部305で解析したアクセス要求に基づいてアービトレーションを行い(ステップ708a、708b)、セレクタ304を切り替える。アクセスパス1:136を通して送られてくるデータはパケットバッファ303に格納する。アービトレーションの結果、メモリモジュール109へのアクセス権を得たら、メモリの制御情報をメモリ制御部307へ送出し、メモリアクセスのための前処理を行う(ステップ710a、710b)。次に、パケットバッファ303からデータを読み

出し、セレクタ304を介してメモリモジュール109へ書き込む(ステップ711a、711b)。

【0083】メモリモジュール109へのアクセスが終了すると、メモリアクセスの後処理を行い、データ転送制御部315においてアクセス状況を示すステータス(STATUS)を生成する(ステップ712a、712b)。次に、ステータスをセレクタ部13を介してCMアクセス制御部104へ送出する(ステップ713a、713b)。セレクタ部13内のデータ転送制御部315はステータスを受け取ると、CMコントローラa、b:107へのREQ信号をそれぞれオフする(ステップ714a、714b)。また、セレクタ部13はCMコントローラa、b:107の両方からステータスを受け取ったら、それらを続けてCMアクセス制御部へ送出する(ステップ715)。CMアクセス制御部104内のデータ転送制御部310は2つのステータスを受け取ると、セレクタ部13へのREQ信号をオフする(ステップ715)。セレクタ部13内のデータ転送制御部315はCMアクセス制御部104からのREQ信号のオフを確認すると、CMアクセス制御部104へのACK信号をオフする(ステップ716)。

【0084】CMアクセス制御部104内のデータ転送制御部310はステータスを受け取ると、制御線1:211により、ホストIF102あるいはドライブIF103へキャッシュメモリ部14へのアクセスの終了を報告する。

【0085】上記のように二重化したキャッシュメモリ部14にデータを二重に書き込む場合、二重に書き込むデータの一方のデータの書き込み処理を行っている間に、もう一方のデータが他のアクセスパスからの書き込み要求によって書き換えられるのを防ぐ必要がある。本実施例では、共有メモリ部15にキャッシュメモリ部14のディレクトリを格納し、キャッシュメモリ部14にアクセスする前に必ず、共有メモリ部15に格納したディレクトリにアクセス中を示すビットを立てる。これにより、キャッシュメモリ部14内の同じアドレスには同時に1つのアクセス要求しか発行されないため、二重に書き込むデータの一方のデータの書き込み処理を行っている間に、もう一方のデータが他のアクセスパスからの書き込み要求によって書き換えられるのを防ぐことができる。

【0086】【実施例3】ディスクアレイ制御装置1では、キャッシュメモリ部14を複数設けた場合、あるキャッシュメモリ部14から別のキャッシュメモリ部14へデータをコピーする機能が要求される。この機能は、以下に述べる手順で実現できる。

【0087】アクセス開始の指示を受けたホストIF102あるいはドライブIF103は、制御線1:211によりCMアクセス制御部104内のデータ転送制御部310へアクセス開始を示す信号を送出する。それとと

もに、データ線210を通して2つのアドレス、コマンドを送出する。2つのうちの1つのアドレス及びコマンドは、コピー元のアドレスとリードコマンドで、もう1つのアドレス及びコマンドはコピー先のアドレスとライトコマンドである。ここでは、CMコントローラa:107をコピー元、CMコントローラbをコピー先として説明する。

【0088】CMアクセス制御部104は、データ線210を通して送られてきた2つのアドレス、コマンドをパケットバッファ303に格納する。データ転送制御部310はアービトレーションを行ってバスIF301の使用権を決定し、セレクタ302を切り替える。

【0089】図12は、2つのキャッシュメモリ部間でデータをコピーする場合の、CMアクセス制御部104からCMコントローラa、b:107へのアクセスの流れを示している。CMアクセス制御部104内のデータ転送制御部310は、アービトレーションによってアクセスパス0:135の使用権が決定されると、制御線2:212によってセレクタ部13内のデータ転送制御部315へアクセス開始を示す信号(REQ)を出す(ステップ801)。続いて、アドレス及びコマンドを2つ連続して送出する(ステップ802)。

【0090】セレクタ部13内のデータ転送制御部315は、CMアクセス制御部104からREQ信号を受け取ると、次にアクセスパス0:135を通して送られてくるアドレス及びコマンドを受信し、adr、cmd解析部305で解析したアクセス要求に基づいてアービトレーションを行う(ステップ803)。アービトレーションの結果、CMコントローラa、b:107への2つのアクセスパス1:136への接続権の両方を得たら、データ転送制御部315はセレクタ306を切り替える(ステップ804)とともに、制御線2:212により、CMアクセス制御部104内のデータ転送制御部310へ、CMコントローラa、b:107両方への接続権が得られたことを示す信号(ACK)を返す(ステップ805)。次にデータ転送制御部315は、制御線3:213によってCMコントローラa、b:107内のデータ転送制御部315へアクセス開始を示す信号(REQ)を出す(ステップ806a、806b)。続いて、CMコントローラa、b:107へそれぞれのアドレス及びコマンドを送出する(ステップ807a、807b)。

【0091】CMアクセス制御部104はACK信号を受けると、アクセスの終了を知らせるステータスの受信待ち状態に入る。

【0092】コピー元のCMコントローラa:107内のデータ転送制御部315は、制御線3:213によってREQ信号を受け取ると、次にアクセスパス1:136を通して送られてくるアドレス及びコマンドを受信し、adr、cmd解析部305で解析したアクセス要

求に基づいてアービトレーションを行い(ステップ808)、セレクタ304を切り替える。アービトレーションの結果、メモリモジュール109へのアクセス権を得たら、メモリの制御情報をメモリ制御部307へ送出し、メモリアクセスのための前処理を行う(ステップ809)。次に、メモリモジュール109からデータを読み出し(ステップ810)、セレクタ304を介してアクセスパス1:136へ送出する(ステップ811a)。

【0093】セレクタ部13は、アクセスパス1:136を通してCMコントローラa:107から送られてきたデータをCMコントローラbに繋がるアクセスパス1:136へ送出する。(ステップ811b)。

【0094】CMコントローラb:107内のデータ転送制御部315は、制御線3:213によってREQ信号を受け取ると、次にアクセスパス1:136を通して送られてくるアドレス及びコマンドを受信し、adr、cmd解析部305で解析する。その後、データを受信し始めるのを待って、メモリアクセスのアービトレーションに参加する(ステップ812)。アクセスパス1:136を通して送られてくるデータはパケットバッファ303に格納する。アービトレーションの結果、メモリモジュール109へのアクセス権を得たら、メモリの制御情報をメモリ制御部307へ送出し、メモリアクセスのための前処理を行う(ステップ813)。次に、パケットバッファ303からデータを読み出し、セレクタ304を介してメモリモジュール109へ書き込む(ステップ814)。

【0095】CMコントローラa、b:107は、それぞれCMメモリモジュール109へのアクセスが終了すると、メモリアクセスの後処理を行い、データ転送制御部315においてアクセス状況を示すステータス(STATUS)を生成する(ステップ815、818)。次に、ステータスをセレクタ部13へ送出する(ステップ816、819)。

【0096】セレクタ部13内のデータ転送制御部315はステータスを受け取ると、CMコントローラa、b:107へのREQ信号をそれぞれオフする(ステップ817、821)。また、セレクタ部13は2つのCMコントローラa、b:107の両方からステータスを受け取ったら、それらを続けてCMアクセス制御部へ送出する(ステップ820)。CMアクセス制御部104内のデータ転送制御部310は2つのステータスを受け取ると、セレクタ部13へのREQ信号をオフする(ステップ822)。セレクタ部13内のデータ転送制御部315はCMアクセス制御部104からのREQ信号のオフを確認すると、CMアクセス制御部104へのACK信号をオフする(ステップ823)。

【0097】CMアクセス制御部104内のデータ転送制御部310はステータスを受け取ると、制御線1:2

11により、ホストIF102あるいはドライブIF103へキャッシュメモリ部14へのアクセスの終了を報告する。

【0098】ディスクアレイ制御装置1ではまた、1つのキャッシュメモリ部14内のあるアドレスから別のアドレスへデータをコピーする機能も要求される。

【0099】この機能は、図9で示したデータの書き込み時の手順において、ステップ511のメモリモジュール109へのライトアクセスの代わりに、メモリモジュール109からデータを読み出してCMコントローラ107内のパケットバッファ303に格納し、続けてそのデータをメモリモジュール109へ書き込むという処理を行うことによって実現できる。

【0100】[実施例4]図1において、2つの共有メモリ部15間で、メモリ領域の全部または、一部を二重化し、共有メモリ部15へのデータの書き込み時に二重化した2つの領域に同じデータを書き込むことにより、データの信頼性を上げることが可能となる。

【0101】二重化した2つの共有メモリ部15へデータを書き込む場合の手順は、以下のようになる。

【0102】マイクロプロセッサ101は、制御線5:221によりSMアクセス制御部105内のデータ転送制御部310へアクセス開始を示す信号を送出する。それとともに、データ線220を通して2つのアドレス、コマンドと、1つのデータを送出する。

【0103】SMアクセス制御部105は、データ線220を通して送られてきた2つのアドレス、コマンドと、1つのデータをパケットバッファ303に格納する。データ転送制御部310はアービトレーションを行ってバスIF301の使用権を決定し、セレクタ302を切り替える。

【0104】図13は、二重化した2つの共有メモリ部15へデータを書き込む場合の、SMアクセス制御部105から2つのSMコントローラ108へのアクセスの流れを示している。2つの共有メモリ部を二重化する場合、一方をマスタ、もう一方をスレーブに設定する。SMアクセス制御部105内のデータ転送制御部310は、アービトレーションによって、まずマスタ側のSMコントローラ108へのアクセスパス2:137の使用権を決定し、制御線6:222によってマスタ側のSMコントローラ108内のデータ転送制御部315へアクセス開始を示す信号(REQ)を出す(ステップ901)。続いて、アドレス、コマンド、及びデータを連続して送出する(ステップ902)。

【0105】マスタ側のSMコントローラ108内のデータ転送制御部315は、制御線6:222によってREQ信号を受け取ると、次にアクセスパス2:137を通して送られてくるアドレス、コマンド、及びデータを受信し、addr、cmd解析部305で解析したアクセス要求に基づいてアービトレーションを行い(ステップ

903)、セレクタ309を切り替える。データはパケットバッファ303に格納する。アービトレーションの結果、メモリモジュール109へのアクセス権を得たら、メモリの制御情報をメモリ制御部307へ送出し、メモリアクセスのための前処理を行う(ステップ904)。次に、パケットバッファ303からデータを読み出し、セレクタ309を介してメモリモジュール109へ書き込む(ステップ905)。

【0106】メモリモジュール109へのアクセスが終了すると、メモリアクセスの後処理を行い、データ転送制御部315においてアクセス状況を示すステータス(STATUS)を生成する(ステップ906)。次に、ステータスをSMアクセス制御部105へ送出する(ステップ907)。

【0107】SMアクセス制御部105はステータスを受け取ったら、マスタ側のSMコントローラ108へのアクセスパス2:137の使用権を開放せずに、アービトレーションによって、スレーブ側のSMコントローラ108へのアクセスパス2:137の使用権を決定する。その後のスレーブ側のSMコントローラ108へのアクセス手順(ステップ908~914)は、マスタ側のSMコントローラ108へのアクセス手順(ステップ901~907)と同様である。

【0108】SMアクセス制御部105内のデータ転送制御部310はスレーブ側のSMコントローラ108からステータスを受け取ると、マスタ側とスレーブ側両方のSMコントローラ108へのREQ信号をオフする(ステップ715a、715b)。

【0109】二重化した共有メモリ部15にデータを二重に書き込む場合、上記のように共有メモリ部15をマスタとスレーブに分け、マスタ側、スレーブ側の順でデータを書き込み、スレーブ側へのデータの書き込みが終わるまで両方のアクセスパス2:137を開放しない。これによりデータを書き込む順番が保証され、二重に書き込むデータの一方のデータの書き込み処理を行っている間に、もう一方のデータが他のアクセスパスからの書き込み要求によって書き換えられるのを防ぐことができる。

【0110】SMアクセス制御部105内のデータ転送制御部310はステータスを受け取ると、制御線5:221により、マイクロプロセッサ101へ共有メモリ部15へのアクセスの終了を報告する。

【0111】[実施例5]図14に、実施例1のディスクアレイ制御装置1のチャネルIF部11、ディスクIF部12、セレクタ部13、キャッシュメモリ部14、共有メモリ部15を実装するときの構成を示す。

【0112】チャネルIF部11、ディスクIF部12、セレクタ部13、キャッシュメモリ部14、共有メモリ部15は、それぞれ独立したパッケージ(PK)、すなわち、チャネルIFPK1、ディスクIFPK2、

セレクタPK3、メモリPK4にそれぞれ実装する。異なるキャッシュメモリ部14は、異なるパッケージ上に実装する。また、異なる共有メモリ部15も、異なるパッケージ上に実装する。1つのキャッシュメモリ部14と1つの共有メモリ部15は同一のパッケージ上に実装しても問題ない。図14では、1つのキャッシュメモリ部14と1つの共有メモリ部15を同一のメモリPK4上に実装した例を示している。

【0113】チャネルIFPK1、ディスクIFPK2、セレクタPK3、メモリPK4は、プラッタ5上に実装し、それを筐体に搭載する。

【0114】ここで、各PKをプラッタ5に実装するときには、セレクタPK3を中心付近より外側、好ましくはプラッタの両端に配置することが重要となる。このような配置により、各PK間を結ぶ線をプラッタ5上に配線する際、プラッタ5全体にわたって配線の密度を均一にでき、プラッタ上の配線を容易になる。

【0115】本実施例では、セレクタ部13をセレクタPK3に実装して、プラッタ5の両端に配置するとした。しかし、セレクタ部13をパッケージに実装せず、プラッタの両端に直接実装しても問題ない。

【0116】【実施例6】図1に示す実施例1のディスクアレイ制御装置1において、SMアクセス制御部105とSMコントローラ108間をアクセスバス2:137で接続する代わりに、図7に示すように、2本の共有バス130を介して接続する。SMアクセス制御部105からは、2本の共有バス130それぞれに1本ずつ接続バスを設ける。また、SMコントローラ108からも、2本の共有バス130それぞれに1本ずつ接続バスを設ける。SMアクセス制御部105及びSMコントローラ108は、それぞれ2つのアービタを有している。2つのアービタはそれぞれ2本の共有バスのアービトリエーション用のアービタである。SMアクセス制御部105からSMコントローラ108へのアクセスの際は、複数のアービタの内の1つがマスターとなり、共有バス130の使用権のアービトリエーションを行う。そして、使用権を得たSMアクセス制御部105がSMコントローラ108にアクセスを行う。また、アービタをSMアクセス制御部105及びSMコントローラ108内に設ける代わりに、独立した回路として共有バス130に直接接続しても問題ない。

【0117】共有バス接続では、バスのデータ幅を広げることによりデータ転送速度を上げることが可能であり、共有メモリ部15へのアクセス時間を短縮可能である。

【0118】上述したように、図1におけるアクセスバス2:137のデータ幅は、アクセスバス0:135のデータ幅よりも2倍以上小さくすることができるので、図1のようにチャネルIF部11及びディスクIF部12と共有メモリ部15との間をスター接続（1対1接

続）しても、共有メモリ部を実装するLSIのピンネックの問題が生じるケースは少ない。とはいっても、アクセスバス2:137の本数が増え過ぎでアクセスバス2:137を実装できないという問題が生じる可能性もある。そうした場合、本実施例の共有バス接続が有効となる。

【0119】なお、本実施例では、耐障害性の向上という観点から、セレクタ部13、キャッシュメモリ部14、及び共有メモリ部15をそれぞれ二重化しているが、これらを二重化しなくても、上述の効果を得られるることは言うまでもない。

【0120】

【発明の効果】本発明によれば、チャネルIF部、ディスクIF部-キャッシュメモリ間のアクセスバスについてはスループットを高くできる。また、チャネルIF部、ディスクIF部-共有メモリ間のアクセスバスについてはスループットを高くでき、かつアクセス時間を短くできる。これによって、スループットが高く、かつ応答時間の短いディスクアレイ制御装置を提供することが可能となる。

【図面の簡単な説明】

【図1】本発明によるディスクアレイ制御装置の構成を示す図。

【図2】従来のディスクアレイ制御装置の構成を示す図。

【図3】従来のディスクアレイ制御装置の他の構成を示す図。

【図4】本発明によるディスクアレイ制御装置内のCMアクセス制御部の構成を示す図。

【図5】本発明によるディスクアレイ制御装置内のSMアクセス制御部の構成を示す図。

【図6】本発明によるディスクアレイ制御装置内のセレクタ部の構成を示す図。

【図7】本発明によるディスクアレイ制御装置内のキャッシュメモリ部の構成を示す図。

【図8】本発明によるディスクアレイ制御装置内の共有メモリ部の構成を示す図。

【図9】キャッシュメモリ部へのデータの書き込み時の手順を示す図。

【図10】共有メモリ部へのデータの書き込み時の手順を示す図。

【図11】二重化した2つのキャッシュメモリ部へデータを二重に書き込む時の手順を示す図。

【図12】1つのキャッシュメモリ部から別のキャッシュメモリ部へデータをコピーする時の手順を示す図。

【図13】二重化した2つの共有メモリ部へデータを二重に書き込む時の手順を示す図。

【図14】本発明によるディスクアレイ制御装置内の実装の構成を示す図。

【図15】本発明によるディスクアレイ制御装置内の他の構成を示す図。

### 【図16】本発明によるディスクアレイ制御装置内の他の構成を示す図。

【図17】本発明によるディスクアレイ制御装置の構成を示す図。

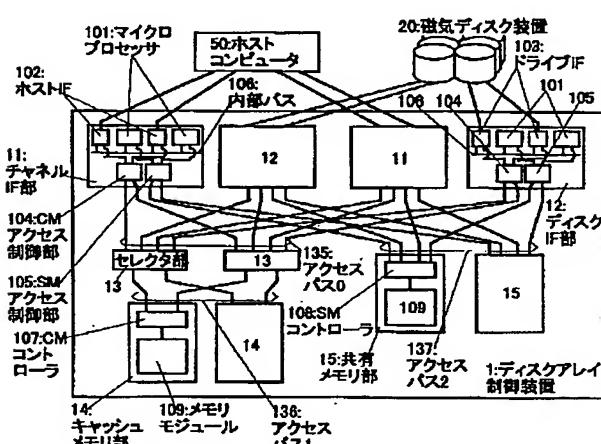
### 【符号の説明】

1…ディスクアレイ制御装置、11…チャネルIF部、  
12…ディスクIF部、13…セレクタ部、14…キャ  
ッシュメモリ部、15…共有メモリ部、20…磁気ディ

スク装置、50…ホストコンピュータ、101…マイクロプロセッサ、102…ホストIF、103…ドライブIF、104…CMアクセス制御部、105…SMアクセス制御部、106…内部バス、107…CMコントローラ、108…SMコントローラ、109…メモリモジュール、135…アクセスバス0、136…アクセスバス1、137…アクセスバス2。

【図1】

圖 1

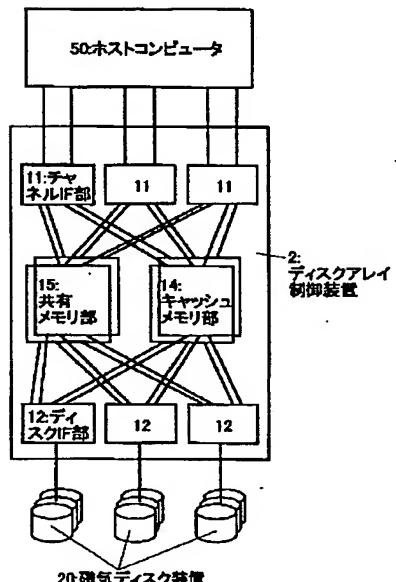


[図3]

3

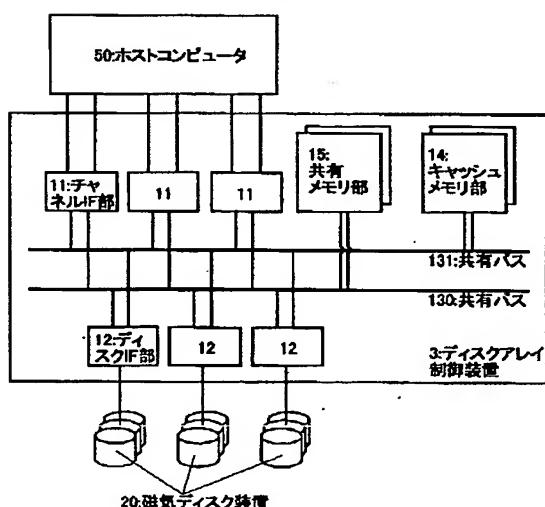
[図2]

図2



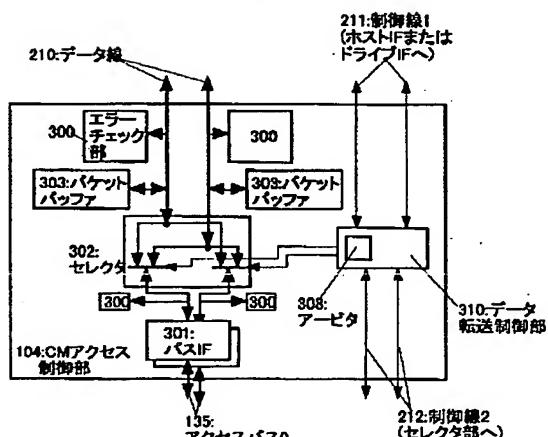
[図4]

圖4



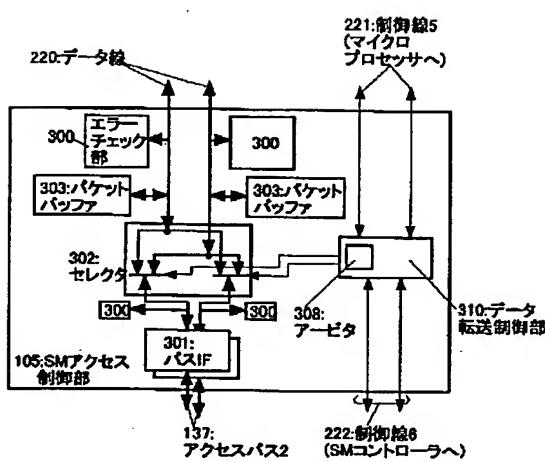
【図4】

図4



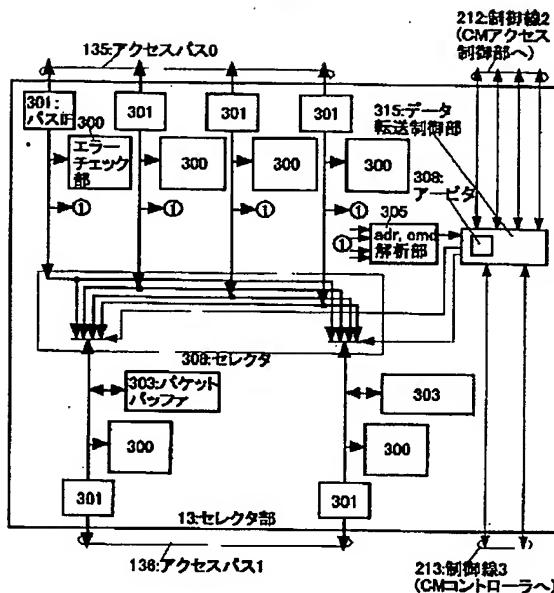
【図5】

圖5



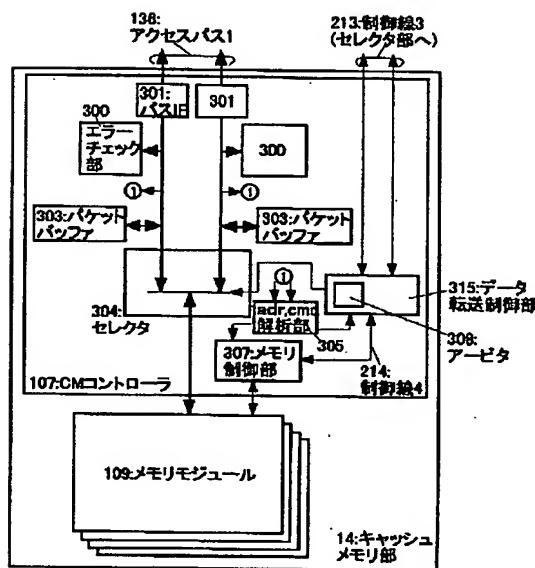
[图 6]

圖 6



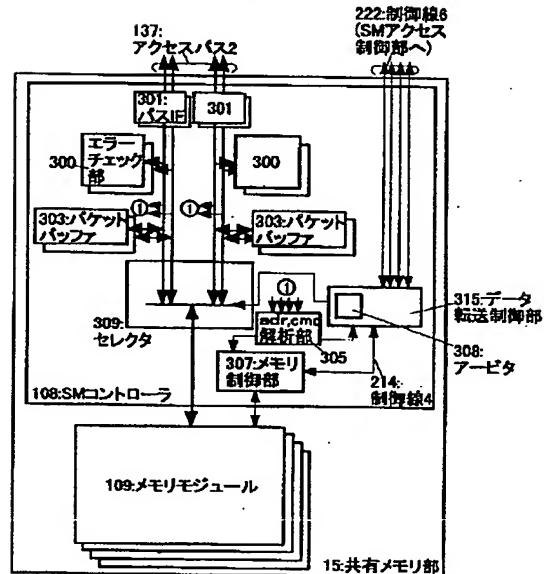
【図7】

圖 7



【図8】

8



【図9】

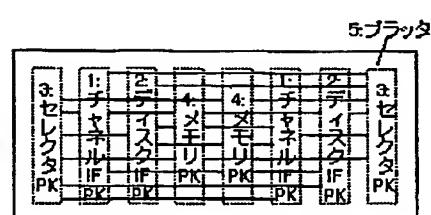
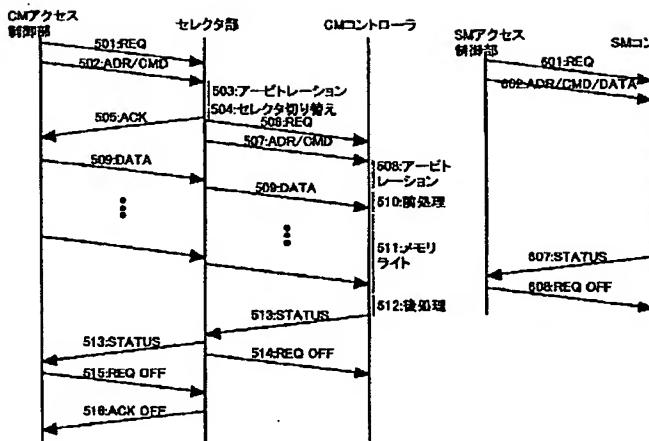
【図10】

【図14】

図9

図10

図14

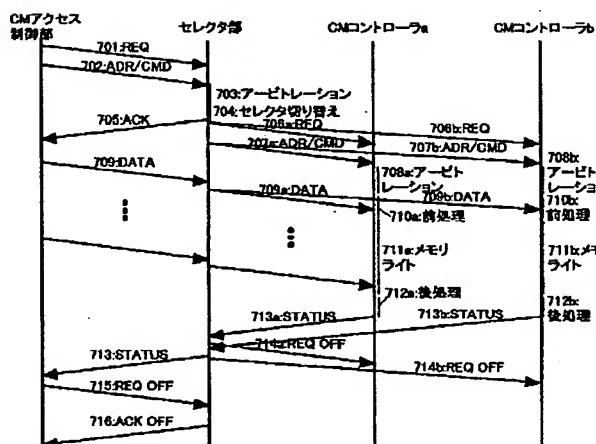
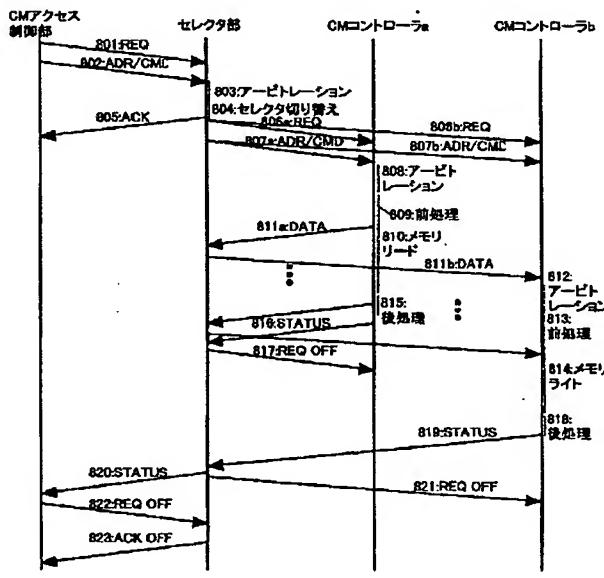


【図12】

図12

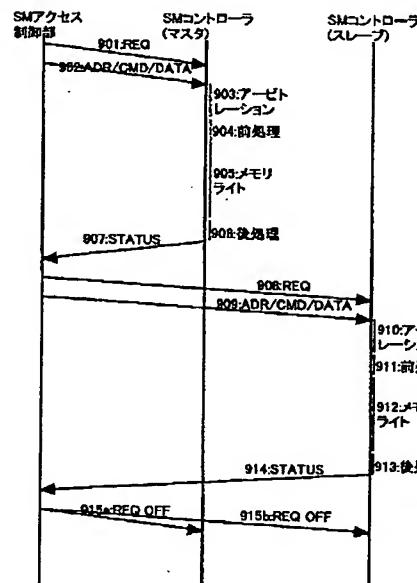
【図11】

図11



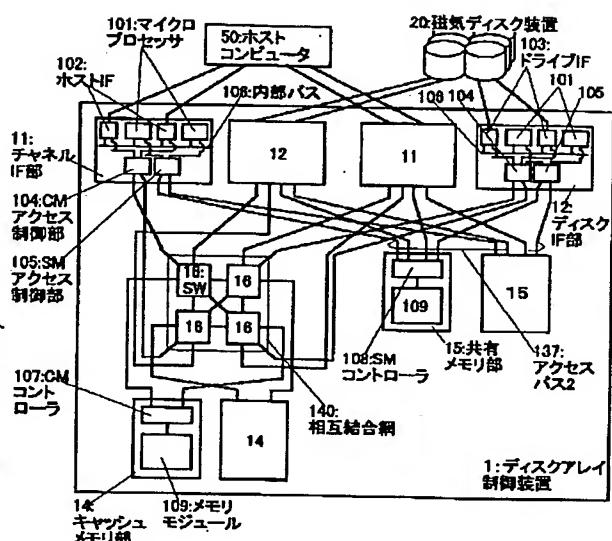
### 【図13】

图 13



【図15】

圖15

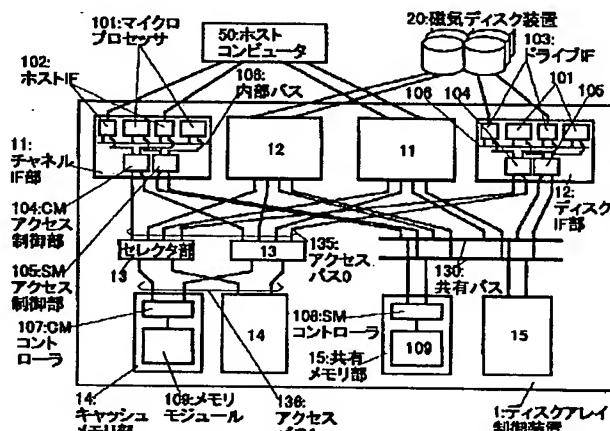
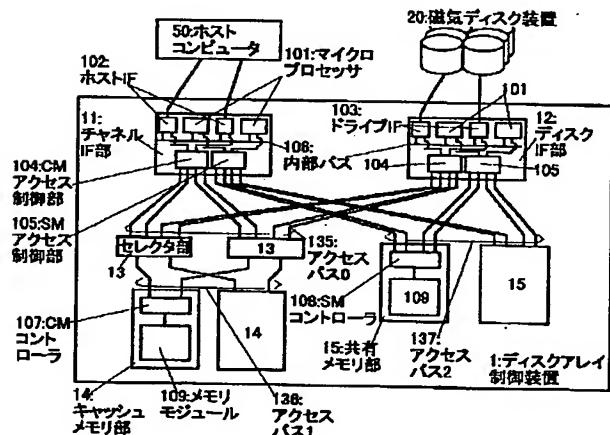


【図17】

【図16】

図17

图16



## フロントページの続き

(72) 発明者 藤林 昭

東京都国分寺市東恋ヶ窪一丁目280番地  
株式会社日立製作所中央研究所内

(72) 発明者 金井 宏樹

東京都国分寺市東恋ヶ窪一丁目280番地  
株式会社日立製作所中央研究所内

(72)発明者 箕輪 信幸

神奈川県小田原市国府津2880番地 株式会

社日立製作所ストレージシステム事業部内

Fターム(参考) 5B065 BA01 CA12 CA30 CE11 CH01